

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 15, 2000, 03:50:49 ; Search time 61.22 Seconds  
(without alignments)  
239.492 Million cell updates/sec

Title: US-09-036-614A-1

Perfect score: 3161  
Sequence: 1 MSGVLGQDEPAGRLSQE.....APLQVSRGLSASTMDLSSSS 619

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 188963 seqs, 23686106 residues

Total number of hits satisfying chosen parameters: 188963

Minimum DB seq length: 0

Maximum DB seq length: 100000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_36.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2000	63.3	569	1 Y07045	Breast cancer asso
2	384	12.1	79	1 Y11521	Human 5' EST seque
3	167.5	5.3	2101	1 R4173	Sequence of the in
4	166	5.3	1972	1 W00024	Smooth muscle myos
5	163	5.2	981	1 W48499	Amino acid sequenc
6	161.5	5.1	2192	1 W21732	LexA/NUMA fusion p
7	161.5	5.1	2272	1 W21731	GAL4/HA/NUMA fusio
8	160	5.1	885	1 R66930	AMML chromosome in
9	159.5	5.0	543	1 Y07058	Renal cancer assoc
10	152.5	4.8	1325	1 W19540	Male-enhanced anti
11	152.5	4.8	1325	1 W94391	Mouse male enhance
12	151.5	4.8	920	1 W82500	Human OGT protein.
13	149.5	4.7	955	1 R57365	K39 polypeptide of
14	149.5	4.7	955	1 W03691	Leishmania chagasi
15	149.5	4.7	1805	1 R27204	Rat nestin. Diagn
16	149.5	4.7	1805	1 R60126	Rat nestin protein
17	148	4.7	2482	1 R72826	Human mitotin. Pur
18	148	4.7	2482	1 W23996	Human mitotin amin
19	148	4.7	3248	1 R99795	Kinetochores protei
20	147	4.7	1120	1 W81642	Mouse elf protein.
21	147	4.7	1151	1 W82501	C. elegans OGT pro
22	146.5	4.6	914	1 W18030	Huntingtin interac
23	146.5	4.6	2954	1 Y01632	Amino acid sequenc
24	145	4.6	2154	1 W81639	Mouse elf-1 protei
25	144	4.6	630	1 W39166	Mouse RHAMM protei
26	144	4.6	816	1 R66931	AMML chromosome in
27	142.5	4.5	977	1 W59881	Amino acid sequenc
28	142	4.5	1027	1 W20217	H. pylori cytoplasm
29	142	4.5	1411	1 W02258	Nucleolar/endosoma
30	142	4.5	2440	1 W20828	H. pylori cytoplas
31	141.5	4.5	606	1 R99673	Receptor for hyalu
32	141.5	4.5	631	1 R99675	RHAMM 1-2a isoform
33	139.5	4.4	476	1 R43563	Hyaluronan recepto

34	139	4.4	1886	1 W54241	Rattus norvegicus
35	137	4.3	546	1 W11214	Leishmania major M
36	137	4.3	546	1 W70209	Leishmania antigen
37	137	4.3	1484	1 W89721	Canine ribosome re
38	136	4.3	576	1 R66929	AMML chromosome in
39	135.5	4.3	1091	1 R70982	Megakaryocyte stem
40	135.5	4.3	1091	1 R70983	Megakaryocyte stem
41	135.5	4.3	1354	1 W23654	Physiologically ac
42	135.5	4.3	1354	1 W71020	A modified Rho tar
43	135.5	4.3	1354	1 Y07082	Renal cancer assoc
44	133.5	4.2	945	1 W73624	Human secreted pro
45	132	4.2	1427	1 R10534	Human 160kD mediat

ALIGNMENTS

RESULT 1	
Y07045	
ID	Y07045 standard; Protein; 569 AA.
AC	Y07045:
DT	02-JUL-1999 (first entry)
DE	Breast cancer associated antigen precursor sequence.
KW	Cancer associated antigen; diagnosis; research; treatment; human;
KW	breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
KW	prostate cancer.
OS	Homo sapiens.
PN	W0904255-M2.
PD	28-JAN-1999.
PF	13-JUL-1998; U14679.
PR	22-JUN-1998; US-102322.
PR	17-JUL-1997; US-896164.
PR	10-OCT-1997; US-061599.
PR	10-OCT-1997; US-061765.
PR	10-OCT-1997; US-948705.
PR	11-OCT-1997; GB-021697.
PA	(LUDW-) LODWIG INST CANCER RES.
PI	Chen Y, Gout I, Gure A, OHare M, Obata Y, Old LJ,
PI	Pfreundschuh M, Sahin U, Scanlan MJ, Stockert E,
PI	Tureci O;
DR	WPI; 99-132448/11.
PT	New isolated cancer associated nucleic acids and polypeptides -
PT	isolated using sera from cancer patients, used to develop products
PT	for the diagnosis, monitoring or treatment of cancers
PS	Disclosure; Page 429-430; 787pp; English.
CC	The invention relates to a method for diagnosing a disorder characterised
CC	by expression of a human cancer associated antigen precursor coded for by
CC	a nucleic acid molecule (NAM). The method comprises: (a) contacting a
CC	biological sample isolated from a subject with an agent that specifically
CC	binds to the NAM, an expression product or a fragment of an expression
CC	product complexed with an HLA molecule; and (b) determining the
CC	interaction between the agent and the NAM or the expression product as a
CC	determination of the disorder. The products and methods can be used in
CC	the diagnosis, monitoring, research, or treatment of conditions
CC	characterised by the expression of various cancer associated antigens.
CC	The invention provides nucleic acid sequences and encoded polypeptides
CC	which are cancer associated antigen precursors expressed in human breast
CC	cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
CC	lung cancer.
SQ	Sequence 569 AA;

Query Match	63.3%;	Score 2000;	DB 1;	Length 569;
Best Local Similarity	69.8%;	Pred. No. 8.5e-165;		
Matches	395;	Conservative 77;	Mismatches 76;	Indels 18; Gaps 8;
QY	1	MSGVLGQDEPAGRLSQE	ILGSLTRVLSQGLEALRSEHOAVLSQTTTCQQGHE	60
DB	1	MSTMTVKEDEKL--EKLTDQETISKTQVIOGLEALKNEHNSILQSLLETKLKK--DDE	57	
QY	61	EGLVHEKARLRSMENIELGLSEAQVMLALASHLSTVESEKOKLRAQVRLCQENQWLR	120	
DB	58	SNLVEKSNWRKSLNLEMLGLSEAQVMMALSNHNLNNAVESEKOKLRAQVRLCQENQWLR	117	

Wed Aug 16 09:11:46 2000

CC the signal peptide can be used for directing extracellular secretion of  
CC a polypeptide or the insertion of a polypeptide into a membrane, or  
CC importing a polypeptide into a cell.  
SQ Sequence 79 AA;

Query Match 12.1%; Score 384; DB 1; Length 79;  
Best Local Similarity 98.7%; Pred. No. 2.9e-26;  
Matches 76; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 88 MLALASHLSTVESEKQKLRQVRLRCQENQWLRLDELQAGTQRLQSRQVAAQLEEKHL 147  
Db 1 MLALASHLSTVESEKQKLRQVRLRCQENQWLRLDELQAGTQRLQSRQVAAQLEEKHL 60

QY 148 EFLGRLQYDEDEGHTSE 164  
Db 61 EFLGRLQYDEDEGHTSE 77

RESULT 3  
R47173  
ID R47173 standard; Protein; 2101 AA.  
AC R47173;  
DT 15-JUL-1994 (first entry)  
DE Sequence of the inner nuclear matrix (INM) protein MT2.  
KW Malignant cell type marker; interior nuclear matrix; MT2; NUMA.  
OS Homo sapiens.  
PN W09400573-A.  
PD 06-JAN-1994.  
PF 21-JUN-1993; U06160.  
PR 22-JUN-1992; US-901701.  
PA (MAYR-) MATRI-TECH INC.  
PI Lidgard GP, Foukati G;  
PI WPI; 94-026210/03.  
DR N-PSDB; Q54841.  
PT Novel malignant cell type markers of the interior nuclear matrix  
PT - used for detecting abnormal cell types e.g. malignant breast,  
PT prostate, lung, etc., for determining deg. of cell death in  
PT tissue, etc.  
PS Claim 39; Page 56-66; 93pp; English.  
CC MT1 and MT2 proteins were isolated from malignant cells essentially  
CC following the method of Penman and Fey, described in US Pat. Nos. 4,  
CC 882,268 and 4,885,236. DNA sequences encoding these proteins were  
CC cloned by probing expression libraries using monoclonal antibodies  
CC raised against the isolated INM proteins. The genetic sequence  
CC encoding MT2 has been disclosed by Yang et al. 1992, J. Cell Biol.  
CC 116:1395-1408, and is referred to by them as 'NUMA'. The nucleotide  
CC sequence encoding MT1 has not been described previously. Both the  
CC MT1 and MT2 proteins are distributed throughout the nucleus (with  
CC the exception of the nucleolus) in non-mitotic cells, and localise  
CC to the spindle during mitosis. A binding protein having a binding  
CC affinity of greater than about 10(9) M(-1) for R47173 is claimed  
CC for use in the manufacture of a medicament.  
SQ Sequence 2101 AA;

Query Match 5.3%; Score 167.5; DB 1; Length 2101;  
Best Local Similarity 20.5%; Pred. No. 2.9e-05;  
Matches 156; Conservative 106; Mismatches 272; Indels 227; Gaps 32;

QY 15 HRLSQEILGSTRVLSOGLEA----LRSEHQVQLSQT-----ECIQ--- 55  
Db 750 HKRREK-LSEERAGRGLEARLQLQGEAHOAETVRLRLAFAAQAHTAECSECLVK 808

QY 56 -----QGHEE-----GLVHEKARQLRRSMENIELGLSEAQVMA-LASH-----LS 96  
Db 809 EVAARWDGYSDSQEEEAQYGAMFQEQMTUKECEKARQELQAEKVGIESHLSQIS 868

QY 97 TVESEKRLQRAQVRLQCE-----NQWLDELQAGTQRLQSRQVAAQLE----- 141  
Db 869 RQONKLAELHANLARAQQVQKEVRAQKLDLSTLQEKMAATSEKVARLETIVRAGE 928

QY 142 -EKKHLEFLGQ-LRQYDEDEGHTSEEGD---ATKDSLDLFPNEEEDPSNGLSRGQ 196

QY 121 DELAGTQRLQSRQVAAQLEEKHLLEFLGRLQYDEDEGHTSEEGKDATKSLDLDLP 180  
Db 118 DELANTQKQKQSQVAAQLEEKHLLEFLGRLQYDEDEGHTSEEGKDATKSLDLDLP 177

QY 181 NEEEDPSNGLSR--GOGATAAQGGYEIPARLTLNLNVIQYAAQGRYEVAVPLCKQAL 238  
Db 178 N-DEDDPGQGIQQQSHSAAQAQGGYEIPARLTLNLNVIQYAAQGRYEVAVPLCKQAL 236

QY 239 EDLERTSGRHDPVATMLNIALYVRQNKYKEAAHLLNDALSTRESTGLGDPHVAATL 298  
Db 237 EDLEKTSCHDPVATMLNIALYVRQNKYKDAANLLNDALIREKTLGDKDPHVAATL 296

QY 299 NNLAVLYGKRGYKEAEPLQORALEIREKVLGTNHPDVAKQNLNALLCQNGKYEAVER 358  
Db 297 NNLAVLYGKRGYKEAEPLQORALEIREKVLGKDPDVAKQNLNALLCQNGKYEEVEY 356

QY 359 YYQALAIYEGQLGDPNPNVARTKNNLASCYLKQKGYAEATLYKEILTRAHVQFGSVD 418  
Db 357 YYQALAIYQTKLGPDDPNVAKTKNNLASCYLKQKGYAEATLYKEILTRAHREFGSVD 416

QY 419 DDHRPIWHAEREEMSKSRHHEGGTPVAYEGYKACKYSSPTVNTTLNGLALYRQ 478  
Db 417 DENKPIWHAEREEMSKSRHHEGGTPVAYEGYKACKYSSPTVNTTLNGLALYRQ 474

QY 479 KLEAAETLEECALRRRGTDPIQTKVAELLGESDG---RRTSQEGPGDSVKFE----G 531  
Db 475 KFEAAETLEEAAMRSRKGGLNVHVKQVRAEVLNDPENMEKRRSRESLNVVVKYESGPDG 534

QY 532 GEDASVAVVWSDGSGTTLQSGSLGK 557  
Db 535 GEEVSMSVWNGVSG---RASFCGK 557

RESULT 2  
Y11521  
ID Y11521 standard; Protein; 79 AA.  
AC Y11521;  
DT 21-JUN-1999 (first entry)  
DE Human 5' EST secreted protein SEQ ID NO 343.  
KW Human; secreted protein; EST; expressed sequence tag; diagnosis;  
KW forensic; gene therapy; chromosome mapping; signal peptide;  
KW upstream regulatory sequence; cytokine activity; cell proliferation;  
KW differentiation; haematopoiesis regulation; tissue growth regulation;  
KW reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;  
KW thrombolytic; anti-inflammatory; tumour inhibition.  
OS Homo sapiens.  
PN W09306551-A2.  
PD 11-FEB-1999.  
PF 31-JUL-1998; I81235.  
PR 01-AUG-1997; US-905133.  
PA (GEST) GENSET.  
PI Duclert A, Dumas Milne Edwards J, Lacroix B;  
PI WPI; 99-153781/13.  
DR N-PSDB; X33587.  
PT New nucleic acids encoding human secreted - proteins obtained from  
PT cDNA libraries prepared from substantia nigra, cerebellum, surrenals  
PT and fetal brain tissue  
PS Claim 34; Page 424-425; 434pp; English.  
CC X39440 to X39597 represent 5' expressed sequence tags (ESTs) for human  
CC secreted proteins, and encode the proteins given in Y11374 to Y11531,  
CC respectively. The proteins given represent the signal peptide and an  
CC N-terminal fragment of a secreted protein. The nucleic acid sequences  
CC can be used for producing secreted human gene products. They can also  
CC be used to develop products for diagnosis and therapy. The proteins  
CC obtained may have cytokine activity, cell proliferation/differentiation  
CC activity, haematopoiesis regulating activity, tissue growth regulating  
CC activity, reproductive hormone regulating activity, chemotactic/  
CC chemokinetic activity, haemostatic and thrombolytic activity, receptor/  
CC ligand activity, anti-inflammatory activity, tumour inhibition activity  
CC or other activities. The products can be used in forensic, gene therapy  
CC and chromosome mapping procedures. The sequences can also be used for  
CC obtaining corresponding promoter sequences. The nucleic acids encoding

Db 929 QQETASRLVKKPAGRDQPPWLEEQQGRFCSTQAALQAM--EREAEQMGNELERJA 986  
QY 197 ATAAQQGGYEIPARLTLNLVLIQYAAQGRYEVAVPLCKQALEDLERTSGRGHPDVATML 256  
Db 987 ALMESQGO-----QQEERGQ-----REVARLTQERGRAQADLALK 1024  
QY 257 NILA-LVYRDQNKYE-----AAHLNDALSTRESTGLGPDHPAVALTNLANVLVG-K 307  
Db 1025 AARAELEMLRQNALNEQRFVATQALAHALTEKE---GKD-----QELAKRLGLE 1073  
QY 308 RGKYKEAEPPLQORALEIREKV-----LGTNHPDVAK-----QLANLA 344  
Db 1074 RAQIKLELELRTQVKQLEQAKKEHAGSGQASEAAGRTETGPKLEALRAEVSKLE 1133  
QY 345 LLCQNGKY-EAVERYQO-----RALAIYEQO-----LGPDPNPVARTKNLA 386  
Db 1134 QOCQKQEQADSLERSLEAERASRAERDSALETLQGLEEKAQELGHSQSALASAQRELA 1193  
QY 387 SCYLKQGYAEATLYK-----EILTRAHVQFGSVDDHKK 422  
Db 1194 APTRTVQDHSKAEDEWKAQVANGROAERKNLSLISSEEVSIILNRQVLEKEGESKELKR 1253  
QY 423 PIWMAEEREEMSKSRHHEGGTPYAEYGGWKACKVSSPTVNTTLNLGALYRRQGLEA 482  
Db 1254 LVMAESESQKLEES-----CACCRQOPATVPLOLNAALCGRRCRASG 1298  
QY 483 AETLEECALRSRQCTDPISTQKVAELLGSDGRNTSQEGPGDSVKFEGGEDASVAVWS 542  
Db 1299 REA-EKQVASENLQELTQSAERAELG-----QELKAWQEKFPQKEQALSTILOE 1349  
QY 543 GDGSGTL-----QRSGLKIR-DVLRRSSEL--L 569  
Db 1350 HTSTQALVSELLPAKHLCOQLOAEQAAAEKRRHELEQSKQAAGLRAELRAQRELGE 1409  
QY 570 VKLQGTGTPRSSNNKRAASLNLYNQPSAALPQVSRGLSA 610  
Db 1410 IPLRKVAEQETAQOLRAEKASVAEQLSM--LKAHGLLA 1448  
RESULT 4  
W00024  
ID W00024 standard; Protein; 1972 AA.  
AC W00024;  
DT 25-MAR-1997 (first entry)  
DE Smooth muscle myosin heavy chain SM1 isoform protein.  
KW Smooth muscle; myosin heavy chain; SM1 isoform; rabbit; arteriosclerosis;  
KW gene therapy; mouse; SM2 isoform; retrovirus; adenovirus; restenosis;  
KW associated adenovirus; coronary artery catheterisation; sclerotic artery.  
OS Mus musculus.  
PN W09623069-A1.  
PD 01-AUG-1996.  
PF 25-JAN-1996;  
PR 25-JAN-1995; JP-010085.  
PA (OSAP) OSAKA PREFECTURAL GOVERNMENT.  
PA (VSS-) VESSEL RES LAB CO LTD.  
PI Arakawa E, Hasegawa K, Ishiyama H, Matsuda Y, Oda S;  
PI Sugawara M, Takahashi K;  
DR WPI; 96-362693/36.  
DR N-PSDB; T34291.  
PT DNA encoding smooth muscle myosin heavy chain SM1 isoform - used in  
PT a vector for gene therapy for reduction of re-stenosis following  
PT coronary artery catheterisation  
PS Claim 5; Page 14-27; 42pp; Japanese.  
CC This sequence represents the smooth muscle myosin heavy chain SM1 isoform  
CC protein. The SM1 coding sequence was isolated from a mouse embryo cDNA  
CC library using probes based on the rabbit smooth muscle myosin heavy chain  
CC SM2 isoform. The isolated coding sequence is included in the gene  
CC therapy vector of the invention. The vector of the invention consists of  
CC the coding sequence inserted into a retrovirus, adenovirus, associated  
CC adenovirus or animal-expression plasmid vector (such as pCMX2 or  
CC PAGE208). The vector can be used in the gene therapy treatment of  
CC arteriosclerosis, particularly for the reduction of restenosis occurring

CC following coronary artery catheterisation for widening of sclerotic  
CC arteries.  
SQ Sequence 1972 AA;  
Query Match 5.3%; Score 166; DB 1; Length 1972;  
Best Local Similarity 20.7%; Pred. No. 3.5e-05;  
Matches 140; Conservative 120; Mismatches 278; Indels 138; Gaps 25;  
QY 5 VLGORDEPAGHRLSQEILGSTRVLSQGLEALRSEHQAVLSQSLQSTIECLOOQGGHE-EG 63  
Db 1242 VLQCAQVEVHKKKLEV-----QLQDQKSCDGERARAEKLSKVHLQNEVESVTQM 1295  
QY 64 VHE---KARQLRSMENI-----ELGLSEAQVMALASHLSTVSESKQLRAQVRRL 112  
Db 1296 LNEAGKATKAKDVASLGSLQDQTELLQETQKLVNSTKLRQLEDERNISLQDQDDEE 1355  
QY 113 CQENQ-----WRLDELACTQORLQSRQAVALEEKKHL--EFLQOLROYDED--GH 161  
Db 1356 MEAKQNLERHVSNTLQSDSKKLQDFASTIEVMEBEGKRLQKEMEGLSQOYEKAAAY 1415  
QY 162 TSEKEGDATKQDLDLFPN-EEEDPNSGLSRGOGATAAQGGYEIPARLTLNLVYQ 220  
Db 1416 DKLEKTKNRLQELDLVVDLQNRQLVSNLEKKQ-----KKFDQLAEENKISSK 1466  
QY 221 YA-----AQGRYEVAVPLCK-----QALEDLERTSGRGHPDVATMLNLALVYRD 265  
Db 1467 YADERDRAEAAREKETKALSARALEEALEAKEELERTNKMLKAEMEDLVSSKDDVGN 1526  
QY 266 QNKYKAAHLNDALSIRESTLGPDPHVAATLN-----NLAVLYKRGKYKEAEPIC 318  
Db 1527 VHELEKSKRALETQMEEMTKOLESEDDVOATEDAKLRLEVNNQALKGO-----F 1576  
QY 319 ORALETRKVLGTHPDVAKLNNLALLCONOGKYEAVERYQORALAIYEGOLGPDNPV 378  
Db 1577 ERDLQARDE-----QNEEK-----RRQQRQJHEYTELEDERKOR 1612  
QY 379 A-----RTKNLASCYLK-----QGRYAEATLYKEILTRAHVQEFSGVDDHKKPIM 426  
Db 1613 ALAAAAKKLEGDLKDLEQADSATKGREAIKQLRK---LQAKMDQFQRELLDAR--- 1665  
QY 427 HAEEREEMSKSRHHEGGTPYAEYGGWKACKVSSPTVNTTLNLGAL--YRQGLEAAE 484  
Db 1666 -ASRDEIFATSKENE-----KKAKSLEADLMQLEDAALAAERAKQADLEKEE 1712  
QY 485 TLEECAL----RSRRQGTDPISQTKVAELLGSDGRRTSQEGPGDSVKFEGGEDASVAVE 540  
Db 1713 LAELASSLSGRNTLODEKRRLEARIQAEELEEEQGNNEAMSQRVRAKTLQASOLSNE 1772  
QY 541 WSGDGSGLTQSRGSLGKIRDLVRRSELVLRKLGQTGTEPRPSSNNKRAASLN-YNQPSA 599  
Db 1773 LATERS-TAKNES---ARQQLERQNKELSKLQVEGAVKAKLSTVALEAKIAQLEE 1828  
QY 600 APLOVSRGSLASMTDL 615  
Db 1829 QVEQEAEREKQAATKSL 1844  
RESULT 5  
W46499  
ID W46499 standard; Protein; 981 AA.  
AC W46499;  
DT 19-MAY-1998 (first entry)  
DE Amino acid sequence of the spindly (SPY) protein of Arabidopsis.  
KW Gibberellin signal transduction; spindly phenotype; SPY gene; rescue;  
KW spy mutant gene; gibberellin overdose syndrome; paclobutrazol; spy-4 DNA;  
KW modulation; plant development; plant height; fruit growth;  
KW flower development; leaf size.  
OS Arabidopsis thaliana.  
FH Key Location/Qualifiers  
FT Region 1..76  
FT Domain /note= "N-terminus"  
FT 77..430

FT /note= "Tetraco peptide repeat region which is  
FT important for SPY protein function"

FT Region 77. 110  
FT /note= "Tetraco peptide repeat 1"  
FT 111. 151  
FT /note= "Tetraco peptide repeat 2"  
FT 152. 185  
FT /note= "Tetraco peptide repeat 3"  
FT 186. 219  
FT /note= "Tetraco peptide repeat 4"  
FT 220. 260  
FT /note= "Tetraco peptide repeat 5"  
FT 261. 294  
FT /note= "Tetraco peptide repeat 6"  
FT 295. 328  
FT /note= "Tetraco peptide repeat 7"  
FT 329. 362  
FT /note= "Tetraco peptide repeat 8"  
FT 363. 396  
FT /note= "Tetraco peptide repeat 9"  
FT 397. 430  
FT /note= "Tetraco peptide repeat 10"  
FT 431. 914  
FT /note= "C-terminus"

PN W09743419-A2.

PD 20-NOV-1997.

PF 16-MAY-1997; U08765.

PR 16-MAY-1996; US-649046.

PA (MINU ) UNIV MINNESOTA.

PI Jacobsen SE, Olszewski NE;

DR WPI: 98-008688/01.

DR N-PSDB; V05171.

PT New isolated spindly gene from plants - is involved in gibberellin  
PT signal transduction, used to develop products for altering plant  
PT development

PS Disclosure; Fig 2: 54pp; English.

CC The present sequence represents a protein that is involved in gibberellin  
CC signal transduction. Inactivation of the gene produces a spindly  
CC phenotype (hence the SPY gene). The spindly mutation is characterised by  
CC elongated petioles, yellow-green leaves, early flowering, long spindly  
CC bolts, partial male sterility and parthenocarpic fruit development. These  
CC phenotypes are also observed in wild type plants exhibiting a gibberellin  
CC overdose syndrome due to external application of gibberellin. A spy  
CC phenotypic mutant was isolated from a library of Agrobacterium-mediated  
CC seed transformation lines of Arabidopsis having T-DNA insertions  
CC positioned throughout the genome. Seeds from the library were selected  
CC for their ability to germinate in the presence of paclobutrazol.  
CC Restriction mapping was used to identify spy mutant (spy-4) DNA from  
CC resistant seeds. This DNA was used as a probe to obtain the SPY cDNA.  
CC Introduction of the SPY gene into plants rescues the spindly phenotype.  
CC The SPY DNA, vectors and proteins can be used to modulate plant  
CC development including plant height, fruit growth, flower development and  
CC leaf size.  
CC Sequence 981 AA;

Query Match 5.2%; Score 163; DB 1; Length 981;

Best Local Similarity 22.6%; Pred. No. 2.1e-05;

Matches 95; Conservative 67; Mismatches 168; Indels 90; Gaps 18;

QY 214 LHNLIQYAAQGYEYAVPLCKQALDERTSGRHPDVATMLNIALYTRQNKYKEAA 273

DB 365 MYNLGVAHGEMLFDMVAIFYELAF-----HFNPCHAEACNGLVLYKDRNDLRAV 416

QY 274 HLLNDALSIRETLGPDHPVAATLNNLAVLYGKRGYKEAEPLCGORALEIREKVLGTNH 333

DB 417 ECQMALSIK-----PFAQSLNGLGVYTVQKMDAAASWIEKAI-----LTN 460

QY 334 PDVAKQLNNLALICQNGKYVEARYYQALAIYEGQLGPDNPVARTKNNLASCYLKGG 393

DB 461 PTVAEAFNGLVLYRDAGNITMAIDAYECL-----KIDPDSRNAGQNR--LLAMNYINEG 514

QY 394 ---KYAEATLYNEILLTRAHVQ--EFGSVDDHKPI-----WMAEEREENSRSR 439

Db 515 LDDKLFEAHRDWCGRFTRLHPQYTSWDLNKDPERPTIIGYISPDFTHSVSFIAPLTH 574  
QY 440 HEGTPTVAEYGGWKACKYSSPTVNTTLNLGALYRRQGLKLEAAETLEECALRSRRQGT 499  
Db 575 HD-----YTKY-----KVYYSAVVKADAKT---YFRDKV-----LKKGGVWKDIYGD 616  
QY 500 PISQTKVAELLGESDGRRTSQEGSDSVKFEKG-----EDASVAVESWGDGSGT-----L 549  
Db 617 ---EKTIASWVRE--DKIDILVELTGTANNKLGTMACRPAPQVQVTWIGYPNTTGLPTVDY 672  
QY 550 QRSGLGKIRDVLRRSELVLRKLGTEPRPSSNMKRAASLNQPSAAPLOYSRGLS 609  
Db 673 RITDSLADPPDTKQKQVEELVR-----LPD-----CFLCYTPSPGAGVCPPTPAUS 718

RESULT 6

W21732

ID W21732 standard; Protein; 2192 AA.

AC W21732;

DT 01-OCT-1997 (first entry)

DE Lexa/Numa fusion protein.

KW NIP-1; NIP-2; Numa; nuclear mitotic apparatus; Numa interacting protein;

KW cell division; proliferation; antibody; Ab; detection;

KW malignant cell growth.

OS Homo sapiens.

FT Key

FT Location/Qualifiers

FT domain 1..87

FT /label= Lexa\_DNA\_binding\_domain

FT peptide 88..94

FT /label= Polylinker

FT protein 95..2192

FT /label= Residues\_18-2116\_of\_Numa

FT region 285..1784

FT /label= Coiled\_coil\_region

PN W09640917-A1.

PD 19-DEC-1996.

PF 07-JUN-1996; U09504.

PR 07-JUN-1995; US-478408.

PA (UYIA ) UNIV YALE.

PI McPherson SMG, Snyder MP;

DR WPI: 97-077270/07.

DR N-PSDB; 777783

PT New nucleic acid encoding nuclear mitotic appts. interacting

PT proteins - useful for modulating cell division and proliferation and

PT in diagnosis

PS Claim 15; Page 42-50; 78pp; English.

CC The sequences given in W21731-32 represent fusion proteins which contain

CC Numa (nuclear mitotic apparatus). The fusion proteins were used in

CC the identification of Numa interacting proteins (NIP's) (see also

CC W21729-30). Compounds which interfere with the interaction of Numa

CC with a known NIP are used to modulate cell division and/or proliferation.

CC Ab, raised conventionally using NIP-1 or -2 as immunogen, are used to

CC detect NIP (or their complexes) and to block their activity for

CC diagnostic or therapeutic use, e.g. to detect defective Numa or NIP

CC which may be markers for aberrant (including malignant) cell growth

CC (which can also be detected by nucleic acid sequencing). Also where

CC malignancy is related to defects in Numa or NIP, it can be treated by

CC administration of the appropriate functional protein.

CC Sequence 2192 AA;

Query Match 5.1%; Score 161.5; DB 1; Length 2192;

Best Local Similarity 21.0%; Pred. No. 0.0001;

Matches 143; Conservative 93; Mismatches 285; Indels 161; Gaps 28;

QY 33 LEALRSRSHQVQLQSLSQTIE-----CLOGGHSEGLVHEKARQLRRSMENIELG 81

Db 683 LEALEKEAKALETLQLOQLQVANEARSATSQTSQAQREKAELSRVVEELQACVETARQE 742

QY 82 LSEAQVMLASHLSTVESEKQKLRAQVRRICQENQWLRLDELQAGTQORL-----QRS 133

Db 743 QHEAQAVAELE--LQLRSEQOK--ATEKERVQAEKQQLQQLQALKESLKVTKGSLEEK 799



CC The sequences given in W21731-32 represent fusion proteins which contain  
 CC NuMA (nuclear mitotic apparatus). The fusion proteins were used in  
 CC the identification of NuMA interacting proteins (NIP's) (see also  
 CC W21729-30). Compounds which interfere with the interaction of NuMA  
 CC with a known NIP are used to modulate cell division and/or proliferation.  
 CC Ab, raised conventionally using NIP-1 or -2 as immunogen, are used to  
 CC detect NIP (or their complexes) and to block their activity for  
 CC diagnostic or therapeutic use, e.g. to detect defective NuMA or NIP  
 CC which may be markers for aberrant (including malignant) cell growth  
 CC (which can also be detected by nucleic acid sequencing). Also where  
 CC malignancy is related to defects in NuMA or NIP, it can be treated by  
 CC administration of the appropriate functional protein.  
 SQ Sequence 2272 AA;

Query Match 5.1%; Score 161.5; DB 1; Length 2272;  
 Best Local Similarity 21.0%; Pred. No. 0.00011;  
 Matches 143; Conservative 93; Mismatches 285; Indels 161; Gaps 28;

QY 33 LEALRSEHQAVLSQSLQIE-----CLOQGGHEGLVHEKARQLRRSMENIELG 81  
 DB 763 LEALEKRAAKLEILQQQLQVANEARDSAQVSTQAQREKAELSRKVEELQACVETARQE 822  
 QY 82 LSEAQVMLASHLSTVSEKQKLRQVRLCQENQWRLDELACTQORL-----QRS 133  
 DB 823 QHEAQAQVAELE--LQLRSEQQK-ATEKRYAQEKDQQLQEQALQKESLVTKGSLEEK 879  
 QY 134 EQAVAQLEBEKHL-EFLQGLRQVDEQHTSEKEGATKDSLDLFPNEEEDPSNGL- 191  
 DB 880 RRAADALEEQRCISELKAETRSVLEQ-HKRRKE-----LEERAGRGLE 925  
 QY 192 SRQGATAAQGGYFIPAR-----LRTLH-----NLVIOYAA-OGRYE----- 228  
 DB 926 ARLOQLGEARQATEVLRRELAEMAAQHTAESECEQLVKEVAARERYEDSQEEAQYG 985  
 QY 229 -----VAVPLCKQAEDEL-----ERTSG-----RGHPDVATM-LNIALVYRD 265  
 DB 986 AMFQELMTLKECEKARQEQAEKQVAGTESHSELQISQQNELRAELHANLARAQQV 1045  
 QY 266 QNKYKEAAHLNDALSIRESTLGPDPHVAATLNNLAVLYGKRGYKEAEPLCORALEIR 325  
 DB 1046 QKEVRAQKLADDLSTLOEK-----MAATSKVARLETIVRKAGEQQETASRELVKE 1097  
 QY 326 EKVLTGNHPDVAKQLNNLALLCONQGYEAVERYQR-----ALAIYEQ-----L 371  
 DB 1098 PARAGDRQPEWLEEQG-RQFCSTQAALQAMERAEQMGNELERLRAALMESQGOQEEER 1156  
 QY 372 GPDNPVAVR--TKNNLASCYLKQGYAEAE-----TLYKEILTRAHVQE 413  
 DB 1157 GQQEREVARLTQERGAQADLALEKAAAELEMLQNALNEQRYEFATLQALAHATEK 1216  
 QY 414 FGSVDDDHKPIWHAEREEMSKSR-----HHEGGTPYAEYGGWYKACK 457  
 DB 1217 EGKQDELAKLGLAEAAQIKLEELRQTVKQLKEQAKKEHSHSGSGAQSEAAG-----RTE 1273  
 QY 458 VSSPTVNTTLRNLGALYRR-QGKLEAAETLEEC--ALRSRROGTDPISQT-----KVAE 508  
 DB 1274 PTGPKLEALRAEVSKLEQCCQKQEQADSLERSLEAERASRAERDSALETLQGLEEKAQ 1333  
 QY 509 LLGESDGRRTSQEGPDSVKFEGGEDASVAVEWSGD---GSGTLQORSGLKIRDVLRRS 565  
 DB 1334 ELGHSQSALASAQRELAFAETRTKVQDHSKAEDWKQAQVARGQEAERKNSL-----ISSLE 1388  
 QY 566 SELLVKRLQGTETPRPSSNNMKR 587  
 DB 1389 EEVSILNRQVLEKEGESKELR 1410

RESULT 8  
 R66930  
 ID R66930 standard; Protein: 885 AA.  
 AC R66930;  
 DT 01-SEP-1995 (first entry)

QY 134 EQAVAQLEBEKHL-EFLQGLRQVDEQHTSEKEGATKDSLDLFPNEEEDPSNGL- 191  
 DB 800 RRAADALEEQRCISELKAETRSVLEQ-HKRRKE-----LEERAGRGLE 845  
 QY 192 SRQGATAAQGGYFIPAR-----LRTLH-----NLVIOYAA-OGRYE----- 228  
 DB 846 ARLOQLGEARQATEVLRRELAEMAAQHTAESECEQLVKEVAARERYEDSQEEAQYG 905  
 QY 229 -----VAVPLCKQAEDEL-----ERTSG-----RGHPDVATM-LNIALVYRD 265  
 DB 906 AMFQELMTLKECEKARQEQAEKQVAGTESHSELQISQQNELRAELHANLARAQQV 965  
 QY 266 QNKYKEAAHLNDALSIRESTLGPDPHVAATLNNLAVLYGKRGYKEAEPLCORALEIR 325  
 DB 966 QKEVRAQKLADDLSTLOEK-----MAATSKVARLETIVRKAGEQQETASRELVKE 1017  
 QY 326 EKVLTGNHPDVAKQLNNLALLCONQGYEAVERYQR-----ALAIYEQ-----L 371  
 DB 1018 PARAGDRQPEWLEEQG-RQFCSTQAALQAMERAEQMGNELERLRAALMESQGOQEEER 1076  
 QY 372 GPDNPVAVR--TKNNLASCYLKQGYAEAE-----TLYKEILTRAHVQE 413  
 DB 1077 GQQEREVARLTQERGAQADLALEKAAAELEMLQNALNEQRYEFATLQALAHATEK 1136  
 QY 414 FGSVDDDHKPIWHAEREEMSKSR-----HHEGGTPYAEYGGWYKACK 457  
 DB 1137 EGKQDELAKLGLAEAAQIKLEELRQTVKQLKEQAKKEHSHSGSGAQSEAAG---RTE 1193  
 QY 458 VSSPTVNTTLRNLGALYRR-QGKLEAAETLEEC--ALRSRROGTDPISQT-----KVAE 508  
 DB 1194 PTGPKLEALRAEVSKLEQCCQKQEQADSLERSLEAERASRAERDSALETLQGLEEKAQ 1253  
 QY 509 LLGESDGRRTSQEGPDSVKFEGGEDASVAVEWSGD---GSGTLQORSGLKIRDVLRRS 565  
 DB 1254 ELGHSQSALASAQRELAFAETRTKVQDHSKAEDWKQAQVARGQEAERKNSL-----ISSLE 1308  
 QY 566 SELLVKRLQGTETPRPSSNNMKR 587  
 DB 1309 EEVSILNRQVLEKEGESKELR 1330

RESULT 7

W21731  
 ID. W21731 standard; Protein: 2272 AA.  
 AC W21731;  
 DT 01-OCT-1997 (first entry)  
 DE GAL4/HA/NuMA fusion protein.  
 KW NIP-1; NIP-2; NuMA; nuclear mitotic apparatus; NuMA interacting protein;  
 KW cell division; proliferation; antibody; Ab; detection;  
 KW malignant cell growth.  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT Domain 1..147  
 FT /label= GAL4\_DNA\_binding\_domain  
 FT peptide 148..174  
 FT /label= Hemagglutinin epitope  
 FT protein 175..2272  
 FT /label= Residues\_18-2116\_of\_NuMA  
 FT region 365..1864  
 FT /label= Coiled\_coil\_region  
 PN W09640917-A1.  
 PD 19-DEC-1996.  
 PF 07-JUN-1996; U09504.  
 PR 07-JUN-1995; US-478408.  
 PA (UYA ) UNIV XALE  
 PI McPherson SMG, Snyder MP;  
 DR WPI; 97-077270/07.  
 DR N-PSDB; T77782.  
 FT New nucleic acid encoding nuclear mitotic appts. interacting  
 FT proteins - useful for modulating cell division and proliferation and  
 FT in diagnosis  
 PS Claim 14; Page 28-36; 78pp; English.

DE AMML chromosome inv(16) product.  
 KW AMML: acute myelomonocytic leukemia; chromosome-16; inversion;  
 KW inv(16); CBF-beta; CBFβ gene; translocation factor; myosin; MYH11;  
 KW SMHC.  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT peptide 1..164  
 FT /label= CBFβ  
 FT /label= MYH11  
 FT peptide 165..885  
 FT /label= MYH11  
 FT W09504067-A.  
 PN 09-FEB-1995.  
 PD 26-JUL-1994; U08530.  
 PE 29-JUL-1993; US-09869.  
 PR (UNMI) UNIV MICHIGAN.  
 PA (TEXA) UNIV TEXAS SYSTEM.  
 PI Claxton D, Collins FS, Liu P, Siciliano MJ;  
 DR WPI; 95-082178/11.  
 DR N-PSDB: Q84589.  
 PT Novel DNA spanning the pericentric inversion of chromosome 16 -  
 PT for the screening of acute myeloid leukaemia  
 PS Claim 4; Page 34-38; 78pp; English.  
 CC PCR was performed on total cellular RNA from 5 AMML patients having  
 CC a pericentric inversion of chromosome-16, M4Eo subtype. Sequencing  
 CC showed the inv(16) fusion to comprise a sequence from the CBFβ  
 CC gene, encoding a novel transcription factor, and the MYH11 gene,  
 CC encoding smooth muscle myosin heavy chain. In 1 patient, nt 1-492  
 CC of the CBFβ gene were fused to nt 994 of MYH11 (shown in  
 CC Q84589; predicted as sequence in R60930). Probes based on inv(16)  
 CC can be used for diagnosis of AMML.  
 SQ Sequence 885 AA;

Query Match 5.18; Score 160; DB 1; Length 885;  
 Best Local Similarity 20.99; Pred. No. 3.3e-05;  
 Matches 148; Conservative 114; Mismatches 274; Indels 172; Gaps 29;

QY 10 DEPAHRLSQE-----EILGTRLSQG---LEALRSEHQAVLSQ 49  
 DB 154 DDRSHREMEKANKNDKNTLEKENADLAGELRVLGQAQVEHKKKLEAQVQLQS 213  
 QY 50 TIECLOQGGHEGLVHEKARQLRSMENIELGSEAQ---VWLA-----LASHLSTV--- 98  
 DB 214 --KC-SDGERARAEIANDKRVHQLNEVSVGTMLNEAEGKAIKADVASLSQDQTQEL 270  
 QY 99 --ESEKQLR--AQVRLCQENQWLDELACTQORLORSEQAVA---OLDEEKKHLEFL 150  
 DB 271 LOETTRQKLVNSTKQLRLEERNSLQDDDEMEAKQNLERHISTNLQLSDSKKKLQDF 330  
 QY 151 GOLROYDEGHTSEEKGDATKSDLDLFPNEEEDPSNGLSRGQATAAQQGYEIPAR 210  
 DB 331 ASTVEALEEGKKRFQKE-----TENLTQCYEEK-----AAAYDKLEKTKNR 371  
 QY 211 L-RTLNLVIVQAAGRIEVAVPLCKQALEDLERTSGRHPDVATMLNLALVYRDQNKY 269  
 DB 372 LQOELDDLVDLDNQ-----RQLVSNLEKKQRKFDQLAEKNI--SKYADERDR 420  
 QY 270 KEA-----AHLINDALSIRETGLDPDPAVAATLNNLAVLYGKRGYKEAEPL 317  
 DB 421 AEAAREKETKALSARALEEALEAKEE--LETKNMLKAEMEDLVSSKDDYGVKNVHELEK 479  
 QY 318 CORALEIREKVLGTHNPVAKOLN-----NLALLQCNQKY-----EA 355  
 DB 480 SKRALETQMEEMKTOLLEEDLEQASDAKLRLVNMQAL---KGQFERDLQARDEQNEE 536  
 QY 356 VERYTORALAIYEGOLGPDNPVATKNLASCYLK-----OGKYAEA 398  
 DB 537 KRQORQRLHVEYTELEDE-----RNERALAAAKKLEGLDKLELQADSAIKGREBAI 591  
 QY 399 ETLYKEILTRAHVOFGSVDDDHK-----PIWMAEEREEMSKSRHHEGTPYAEYGGWYK 454  
 DB 592 KQLRK---LQAQMKDFQRELEADARASRDEIFATAKENEKKAKSLEAD----- 635

QY 455 ACKVSSPTVNTTLNLGAL--YRQKGLAAETLEECAL-----RSRQGTDPISQTKVAE 508  
 DB 636 -----LMQLQEDLAAARARQAADLEKEELAEELASSLSGRNALQDEKRRLEARIAQ 687  
 QY 509 LLGSDGRRTSQEGPGDSVKFEGGEDASVAVENSGDGLQSRGSLGKTRDVLRRSEL 568  
 DB 688 LEEELPEEQGNMEASDRVRKATQQAOLSNELATERS-TAQKNES---ARQLERQNKKE 743  
 QY 569 LVRKLGQTEPRPSSNNMKRAASLN-YLNQPSAAPLQVSRGLSASTMDL 615  
 DB 744 LRSKLHEMAGVSKFKSTIAALEAKTAQLEEQVEQZAREKQAATKSU 791  
 RESULT 9  
 Y07058  
 ID Y07058 standard; Protein; 543 AA.  
 AC Y07058.  
 DT 02-JUL-1999 (first entry)  
 DE Renal cancer associated antigen precursor sequence.  
 KW Cancer associated antigen; diagnosis; research; treatment; human;  
 KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;  
 KW prostate cancer.  
 OS Homo sapiens.  
 PN W09904265-A2.  
 PD 28-JAN-1999.  
 PR 15-JUL-1998; U14679.  
 PR 22-JUN-1998; US-102322.  
 PR 17-JUL-1997; US-896164.  
 PR 10-OCT-1997; US-061599.  
 PR 10-OCT-1997; US-061765.  
 PR 10-OCT-1997; US-948705.  
 PR 11-OCT-1997; GB-021697.  
 PA (LUDW-) LUDWIG INST CANCER RES.  
 PI Chen Y, Gout I, Gure A, Obata Y, Old LJ,  
 PI Pfeundschtuh M, Sahin U, Scanlan MJ, Stockert E,  
 PI Tureci O;  
 DR WPI; 99-132448/11.  
 PT New isolated cancer associated nucleic acids and polypeptides -  
 PT isolated using sera from cancer patients, used to develop products  
 PT for the diagnosis, monitoring or treatment of cancers  
 PS Disclosure: Page 448-449; 787pp; English.  
 CC The invention relates to a method for diagnosing a disorder characterised  
 CC by expression of a human cancer associated antigen precursor coded for by  
 CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a  
 CC biological sample isolated from a subject with an agent that specifically  
 CC binds to the NAM, an expression product or a fragment of an expression  
 CC product complexed with an HLA molecule; and (b) determining the  
 CC interaction between the agent and the NAM or the expression product as a  
 CC determination of the disorder. The products and methods can be used in  
 CC the diagnosis, monitoring, research, or treatment of conditions  
 CC characterised by the expression of various cancer associated antigens.  
 CC The invention provides nucleic acid sequences and encoded polypeptides  
 CC which are cancer associated antigen precursors expressed in human breast  
 CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and  
 CC lung cancer.  
 SQ Sequence 543 AA;

Query Match 5.0%; Score 159.5; DB 1; Length 543;  
 Best Local Similarity 21.4%; Pred. No. 1.7e-05;  
 Matches 86; Conservative 61; Mismatches 137; Indels 117; Gaps 17;

QY 38 SEHQAVLSLSQTIECLOQGGHEGLVHEKAR-QLRRSMENIELGISEAQVMLASHLS 96  
 DB 76 SRKAALAEFLNRFEEAKRT--YEEGLKHEANNPQLKGLQNMNEARLAEKRFM----- 125  
 QY 97 TVESERQKLAQVRRLCQENQWLDELACTQORLORSEQAVAQLEEEKKHLEFLGLQRLQY 156  
 DB 126 -----NPF---NWPNLQKLE-SDPTRTLLSDPTPIRELISQLRKN 162  
 QY 157 DEGHGTS-----BEKGDAT-----KDSLDLDFNPEE--E 185  
 DB 163 PSDLGTKLQDPRIWTTLSVLLGDLGMDDEEEIATPPPPPPKKTKEP--PMEEDLPE 220

```
QY 186 DPSNLSRQ--GATAAQQGGYEI-----PARLRLHNLVIOYAAQGRYEVAVP 232
DB 221 NKKQALKEKELGNDAYKKKDFDTALKHYDKAKELDTNMTYITNQAAVYFEKGYDKCRE 280
QY 233 LCKQALEDLERTSGRHPDVTMLNIALVYRDQNKYKEAAHLLNDALSIRESTLGPDPHP 292
DB 281 LCEKAIE-VGRENREDYQIAKAYARIGNSYFKEKYKDAHFYKNSLAERT---PD-- 334
QY 293 AVAATLNNLAVLYGRGKYKEAPICQRALEI--REKVLGTHNPDVAKQLNLLALLCONQ 350
DB 335 -----VL-----KKCQQAEEKILKEQERLAVINPDLALEKNKNGECFOK 373
QY 351 GXYEAVERYQALAIYEGQLGPDNPVARTKNNLASCYLK 391
DB 374 GDYQAMKHYTEAI-----KRNPKDAKLYSNRAACYTK 406

RESULT 10
ID W19540 standard; Protein; 1325 AA.
AC W19540;
DT 16-SEP-1997 (first entry)
DE Male-enhanced antigen-2.
KW Mouse; MEA-2; detecting mutation.
OS Mus musculus domesticus.
FH key Location/Qualifiers
FT misc_difference 305..320
FT /note= "Not shown in the specification"
PN J09121869-A.
PD 13-MAY-1997.
PF 07-NOV-1995; 311638.
PR 07-NOV-1995; JP-311638.
PA (ITOH-) ITO HAM KK.
DR WPI; 97-314229/29.
DR N-PSDB; T74034.
PT Male-enhanced antigen MEA-2 gene - especially from mouse, useful for
PT detecting mutation(s)
PS Claim 8; Page 9-10; 13pp; Japanese.
CC The present sequence represents male-enhanced antigen-2 (MEA-2), which
CC has been derived from a domestic mouse. The polynucleotide encoding
CC the protein can be used for the detection of mutations affecting the
CC MEA-2 gene.
SQ Sequence 1325 AA;
```

Query Match 4.8%; Score 152.5; DB 1; Length 1325;  
Best Local Similarity 21.3%; Pred. No. 0.00028;  
Matches 134; Conservative 104; Mismatches 219; Indels 173; Gaps 30;

```
QY 7 GORDE---PAGHRLSQEILGSTRLV-----SQGLEALRSEHQAVLSQISQITCLOQGGH 59
DB 513 GEREQLQKVADAAASLEQQLQVKTFLFQDQQLAALQOEHLVDIKLTSTQEAQLQAGQ 572
QY 60 EEGLYHEKARQLRSMENIELG-----LSEAQVMLALASHLSTVESEKQLRAQV 109
DB 573 SLDDLTRYDELQARLEQREADSRDAHFLONEKIVLEVA--LQSAKSKDELDROGA 630
QY 110 RLRCQEN-----QWLRLDELQAGTQORLQSEQAQVLEEKHLL--EFLGQ---LROYD 157
DB 631 RLLEEDTEETSGLLLEQLRDLAVKSNQVHLQQTATLTKQMKQKVEQFVQKQVMEAYR 690
QY 158 EDGHTSEE--KEGDATKDSLDLFPNEEEDPNSLGRSQGATAAQGGYEIP-ARLRTL 214
DB 691 ROATSKDQLINELKATFKRLD-----SEMKELRQELIKLQO---EKKTVVEHSLQKD 741
QY 215 HNLVIOYAAQ-----GRYEVAVPL-----CKQALEDLERTS 245
DB 742 MSLVHQMAELEHGLQSVQKQKDEMEIHLQSLKFDKEQKQMALTEANETLKQIEELOQEA 801
QY 246 GRGHPDVTMLNIALVYRDQNK-----YKEAAHLN---DALSTIRE-----284
DB 802 KKAITEQOKMKRGLSDLSAOKEMKTKKAYENAVSILSRRLQELAKASKEATDAELNQL 861
```

```
QY 285 ---STLGPDPHAPAVATLNNLAVLYGRGKYK---EAE-----PLCQRALETRKVLGTN 332
DB 862 RAQSTGSSDPVLHEKIRALEVELQNVGSKILLELEQELQEVITMTSLEESREKYLE 921
QY 333 HP-----DVAKQLNLLALLCONQGYEAVERYQORALAIYEGOL-GPDNPVARTKNN-- 384
DB 922 DELQESRGFRKIKRL-----ESNKKKLALEHER-----GKLYGLGOSNAALREHNSI 971
QY 385 LASYLKQGGKYABAEATLYKEILTRAHVQBFSGVDDDDHKPIWMHAE---FREEMSKSRHHE 441
DB 972 LETALAKR-----EADLVQLNLQVQAVLQKKEEDROMQLVQALQVLSLEKEMK----- 1021
QY 442 GGPYAEYGGWYKACKVSSPTVNTTLRNLCALYRRQCKLEA-----AETLEBICALRS 493
DB 1022 -----VNSLKEQMAA-----ARIEAGHNRHRHFAATLELSEVKK 1055
QY 494 RROGTDPIQSOTKVAEL--LGESDGRRTSOE 521
DB 1056 ELQAKEHLVOTLQAEVDELQIQDGKH-SQE 1084
```

## RESULT 11

```
W94391
ID W94391 standard; Protein; 1325 AA.
AC W94391;
DT 14-APR-1999 (first entry)
DE Mouse male enhanced antigen 2.
KW Mouse; male enhanced antigen 2; MEA-2; Mus musculus domesticus;
KW spermatogenesis; regulation; contraceptive; sterility; inhibition.
OS Mus sp
PN J11018622-A.
PD 26-JAN-1999.
PF 04-JUL-1997; 179490.
PR 04-JUL-1997; JP-179490.
PA (ITOH-) ITO HAM KK.
DR WPI; 99-160962/14.
DR N-PSDB; X04132.
PT Regulation of spermatogenesis using MEA-2 gene information - using
PT anti-sense oligo- or poly-nucleotide(s), used for production of
PT contraceptive
PS Claim 4; Page 8-12; 27pp; Japanese.
CC The present sequence represents mouse male enhanced antigen 2 (MEA-2).
CC The present invention describes the regulation of spermatogenesis by
CC using MEA-2 information. A non-human living organism can have its
CC spermatogenesis inhibited by breakage of the whole or part of the MEA-2
CC gene. Also described are: (1) the creation of the spermatogenesis-
CC inhibited organism; (2) a drug composition containing an oligonucleotide
CC or polynucleotide containing base sequences that pair with at least part
CC of the MEA-2 gene and are able to inhibit the expression of MEA-2 gene;
CC and (3) the creation of an aimed gene-possessing organism using the
CC spermatogenesis inhibited organism. The organism is useful for producing
CC contraceptive drugs.
SQ Sequence 1325 AA;
```

Query Match 4.8%; Score 152.5; DB 1; Length 1325;  
Best Local Similarity 21.3%; Pred. No. 0.00028;  
Matches 134; Conservative 104; Mismatches 219; Indels 173; Gaps 30;

```
QY 7 GORDE---PAGHRLSQEILGSTRLV-----SQGLEALRSEHQAVLSQISQITCLOQGGH 59
DB 513 GEREQLQKVADAAASLEQQLQVKTFLFQDQQLAALQOEHLVDIKLTSTQEAQLQAGQ 572
QY 60 EEGLYHEKARQLRSMENIELG-----LSEAQVMLALASHLSTVESEKQLRAQV 109
DB 573 SLDDLTRYDELQARLEQREADSRDAHFLONEKIVLEVA--LQSAKSKDELDROGA 630
QY 110 RLRCQEN-----QWLRLDELQAGTQORLQSEQAQVLEEKHLL--EFLGQ---LROYD 157
DB 631 RLLEEDTEETSGLLLEQLRDLAVKSNQVHLQQTATLTKQMKQKVEQFVQKQVMEAYR 690
QY 158 EDGHTSEE--KEGDATKDSLDLFPNEEEDPNSLGRSQGATAAQGGYEIP-ARLRTL 214
```

```

Db 691 RDATSKQOLINELKATKRLD-----SEMKELOELIKLOG-----EKKTVEVHSLQKD 741
Qy 215 HNLVIQYAAQ-----GRYEVAVPL-----CKQALEDLERTS 245
Db 742 MSLVHQMAELHGLQSVQKQERDEMEIHLQSLKFDKQFOMIALTEANETLKQJEEELQQA 801
Qy 246 GRGHPDVATMLNIALVYRQNK-----YEAHLLN-----DALSTRE-----284
Db 802 KKAITEQOKMKRGLSOLTSQAQEMTKTKHAYENAVSILSRLLQELALASKEATDAELNOL 861
Qy 285 ---STLGPDPHVAATLNNLAVLYGRGKYK---EAE-----PLCORALEIREKVLGIN 332
Db 862 RAQSTGSSDPVLHLEKTRALEVLQNVGQSKILLEKELQEVITWTSQEELESREKYLE 921
Qy 333 HP-----DVAKQNLNALLCONQKGYEAVRYORALAIYEGOL-GPDNPNVARTKNN---384
Db 922 DELQESGRFRKIKRL-----EESNKLALALEHER-----GKLTGLGQSNALREHNSI 971
Qy 385 LASCYLKQGYAEATLYKELIPRAHVQEFSGVDDHDKPIWMHAE---BREEMSKSRHHE 441
Db 972 LETALAKR-----EADVLQNLQVAVLQORKEEDQRMKQLVQALQVLSKEKME-----1021
Qy 442 GGTPYAEYGGWKACKVSSPTVNTTLNGLALYRROCKLEA-----AETLEECALRS 493
Db 1022 -----VNSLKQMAA-----ARIEAGHNRHRFKRAITLSEVKK 1055
Qy 494 RQCTDPISOTKYVAEL--LGESDGRRTSQE 521
Db 1056 ELQAKEHLVTLQAEVDELQIQDGKH-SQE 1084

```

## RESULT 12

```

W82500
ID W82500 standard; Protein; 920 AA.
AC W82500;
DT 01-FEB-1999 (first entry)
DE Human OGT protein.
KW OGT; O-linked GlcNAc transferase; uridine; transferase; human; tumour;
KW diphospho-N-acetylglucosamine; polypeptide beta-N-acetylglucosaminyl;
KW prediagnosis; type II diabetes; hyperglycaemia; Alzheimer's disease;
KW metastasis; diagnosis; inhibitor; treatment; diabetes mellitus.
OS Homo sapiens.
PN WO9844123-A2.
PE 08-OCT-1998.
PF 27-MAR-1998; U06101.
PR 31-MAR-1997; US-042270.
PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.
PI Hanover JA, Lubas W;
DR WPI: 98-557118/47.
DR N-PSDB: V69301.
PT Protein exhibiting O-linked GlcNAc transferase activity, OGT -
PT useful, e.g. to assess predisposition to type II diabetes or
PT Alzheimer's or metastatic potential of tumours, and to identify
PT inhibitors
PS Claim 7: Page 31-33: 56pp; English.
CC This sequence represents a novel human O-linked GlcNAc transferase, OGT
CC protein (also known as uridine diphospho-N-acetylglucosamine:
CC polypeptide beta-N-acetylglucosaminyl transferase). This protein is
CC useful to assess predisposition toward type II diabetes in patients
CC suspected of having hyperglycaemia that could evolve into this disease,
CC by assaying OGT activity in red blood cells. It can also be used to
CC assess predisposition toward Alzheimer's disease, to assess the
CC metastatic potential of tumours and to diagnose a tumour with metastatic
CC potential. Ogt can also be used to identify OGT inhibitors, especially in
CC high-throughput assays, useful, e.g. in the treatment of diabetes
CC mellitus, tumour-derived diseases and Alzheimer's disease.
SQ Sequence 920 AA;

```

Query Match 4.8%; Score 151.5; DB 1; Length 920;  
 Best Local Similarity 24.2%; Pred. NO. 0.00019;  
 Matches 80; Conservative 42; Mismatches 112; Indels 97; Gaps 14;

```

Qy 207 IPARLRTLH-----NLVIQYAAQRYEYAVPLCKQALEDLERTSGRHPDVATML 256
Db 119 VAAYLRLSLSPNHAVHGNLACVYVEQGLDLAIDTVRAIE-----LQPHFPDAY 170
Qy 257 NILALVYRDQNKYEAAHLNDALSIRESLGPDHPAPAATLNNLAVLYGRGKYKEAP 316
Db 171 CNLANALKEKGSVAEAEDECYNAL-----RLCPHP-----ADSLNNLANIKREOGNIEEAVR 222
Qy 317 LCORALEIREKVLGTNHPDVAKOLNALLCONQKGYEAVRYORALAI-----366
Db 223 LYRKALEV-----PFEFAAHSNLSLVQOQKLOEALMHYKEAIRISPTFADAYS 274
Qy 367 -----YEGOLG-----PDNPNVARTKNNLASCYLKQKGYAEATLYKEIL-----406
Db 275 MGNTLKEMQDVGALQCVTRAIQINPAFADAHNSLASIHKDSGNIPEIASYRTALKLP 334
Qy 407 -----TRAHVQEFSGVDDHDKPIWMHAE-----EEMSKSR-----HHEGGT 444
Db 335 DFPDAYCNLAHCLQI--VCD-----WTDYDERMKKLVSIIVADQLEKNRPLSVPHHSMY 387
Qy 445 PYAEYGGWKACKVSSPTVNTTLNGLALYR 475
Db 388 PLSH--GERKA--IAERHGNLCLDKINVLHK 414

```

## RESULT 13

```

R57365
ID R57365 standard; Protein; 955 AA.
AC R57365;
DT 08-MAR-1995 (first entry)
DE K39 polypeptide of Leishmania chagasi.
KW Parasite; leishmaniasis; infection; repeat unit; vaccine; antigen;
KW Leishmania chagasi.
OS Leishmania chagasi.
PN WO9416331-A.
PE 21-JUL-1994.
PF 10-JAN-1994; U00324.
PR 15-JAN-1993; US-006676.
PA (IASY-) IASYS CORP.
PI Reed SG;
DR WPI: 94-249402/30.
DR N-PSDB: Q70152.
PT Diagnosis of Leishmaniasis - by determining the presence of
PT antibodies that bind to a K39 repeat unit antigen
PS Disclosure; Page 12-15; 28pp; English.
CC The K39 polypeptide comprises a number of repeated units (described
CC in R57366). Detection of antibodies directed against this repeated
CC unit in a patients sample is indicative of leishmaniasis. The
CC antigenic repeat unit can itself be used as a vaccine to protect
CC against infection by a leishmania parasite.
SQ Sequence 955 AA;

```

Query Match 4.7%; Score 149.5; DB 1; Length 955;  
 Best Local Similarity 20.9%; Pred. NO. 0.0003;  
 Matches 129; Conservative 92; Mismatches 274; Indels 121; Gaps 25;

```

Qy 49 QTIECLQGGGHEGLVH-----EKARQLRRSMENIELGLS--EAQVLMALASHLS 96
Db 378 ETLSTLYASRARDIVNVAQVNEPDPRARRIRELEQMEDMRQAMAGDPPAYVSELKKLA 437
Qy 97 TVESEKOKLRAQVRLCOE---NWLRLDELQAGTQQLRSEQAVQALEEKKHLEFLGQL 153
Db 438 LLESEAKRAADLQALEREREHNVQVRLRLATEAKSELESRAALQE-----EMTATR 492
Qy 154 RQYDE-DGHISEEKEGDTAKDLSLDLFPNNEEDPNSGLSRGCGATAAQOGGYEIPARLR 212
Db 493 RQAKMQALNLRKEEQARKER--ELLKEMAKDAA--LSK-----VRRKKDAEIASERE 543
Qy 213 TLNHLVIQYAAQGR-YEVAVPL-----CKQALEDLERTSGRHPDVATMLNIALVYR 264
Db 544 KLESTVAQLEREQEREVALDQALQTHORKLQEALESSERTAAE-----R 587

```

```

QY 265 DQNYKEAAHLNDALSIRP-----STLGPDPHPAATLNNLAVLYGK----- 307
DQ 588 DQ-----LLOQLTELQSERFQTSQVYTDRELRTRDLQRIQYEGTETELARDVALCAA 639
QY 308 ---RGKYKEA-----EPLQORALE-----IREKVLGTHNPDVAKOLNNLALLCQNOGKYEAV 356
DQ 640 QEMEARHAAVPHLOLLELATEWEDALRERALARERDEAAAEALDAASTSON-ARESAC 698
QY 357 ERYQORALAYEGOLGPDNPVARTKNL-----ASCYLKQCKYAEATLYKEIL-TRAH 410
DQ 699 ER-----LTSLEQQLRESEERAAELASOLEATAAKSSAEQDRESTRATLEQQLRESEAR 753
QY 411 VQFEGSVDDHDKPIWMAHEEREMSKSRHHEGTPYAEYGGYKACKVSSPTVNTTLNL 470
DQ 754 AAELASOLEATAAKSSAEQDRESTRATLEQQLRDSSE-----RAELASOLESTAAKM 808
QY 471 GALLYRROGKLEAAETLEECALRSRQGTDPISO---TKVAELLGESDGRRTS-----Q 520
DQ 809 SA---EQDRESTRATLEQQLRDSSEERAAELASOLESTAAKSSAEQDRESTRATLEQQLR 865
QY 521 EGPDSVKFEGGEDASVAVWMSGSGTQORSGLKIRDVLRRSELLVRKLOGETPRP 580
DQ 866 ESEERAAELASOLESTAAKSSAEQDRESTRATLEQQLRDSSEERAAE-LASOLEATAAK 924
QY 581 SSSNMKRAASLNYLNQ 596
DQ 925 SSAEQDRENTRAALEQ 940

RESULT 14
ID W03691 standard; Protein; 955 AA.
AC W03691;
DT 09-MAR-1997 (first entry)
DE Leishmania chagasi K39 antigen.
KW Leishmania chagasi; acidic ribosomal antigen; LcP0;
KW epitope; K39.
OS Leishmania chagasi.
PN W09633414-A2.
PD 24-OCT-1996.
PF 19-APR-1996; U05472.
PR 21-APR-1995; US-428414.
PA (CORI-) CORIXA CORP.
PI Read SG.
DR WFI; 96-485884/48.
DR N-PSDB; T42166.
PT New Leishmania acidic ribosomal P-protein family poly:peptide - used to develop prods. for diagnosis, detection and protection against Leishmania infections
PS Disclosure; Page 36-43; 76pp; English.
CC Compounds including polypeptides that contain at least an epitope of the L. chagasi acidic ribosomal antigen LcP0 are useful in a variety of immunoassays for detecting Leishmania infection. Portions of LcP0 (T42164) contg. at least the 17 C-terminal amino acids (T42165) have been found to generate a signal in an ELISA that is equivalent to that generated by the full length LcP0. A combination of polypeptide may also be used, comprising an LcP0 epitope along with an epitope derived from the Leishmania K39 antigen (T42166), pref. the K39 repeat unit antigen having the sequence given in W03690.
SQ Sequence 955 AA;

Query Match 4.7%; Score 149.5; DB 1; Length 955;
Best Local Similarity 20.9%; Pred. No. 0.0003;
Matches 129; Conservative 92; Mismatches 274; Indels 121; Gaps 25;

QY 49 QTICLCQOQGHGELVH-----EKARQLRRSMENIELGLS--EAQVLMALASHLS 96
DQ 378 ETLSTLYASRRARDIVNVAQVNDPRARRIRELEQMDRQAMAGGDPAYVSELKKLA 437
QY 97 TVESEKOKLAQVRRLCQE---NOWLRDELQGTQORLQSEQAQOLEEKKHLEFLQGL 153

```

```

Db 438 LLESEAKRAADLQALERERHQNVOERLLRATERAEKSELESRAALQOE-----EMTATR 492
QY 154 ROYDE-DGHTSEKEGDATKOSLDLFFNEEEDPNSGLSRGOGATAAQOGGYEIPARLR 212
DQ 493 ROADKQALNLRKEEQARKER--ELLKEMAKDAA--LSK-----VRRRKDABIASERE 543
QY 213 TTHNLVIOYAAQGR-YEYAVPL-----CKQALEDLERTSGRGHPDVATMLNLALVYR 264
DQ 544 KLESTVAQLEREQREVALDALQTHQKLOEALLESSERTAAE-----R 587
QY 265 DQNYKEAAHLNDALSIRP-----STLGPDPHPAATLNNLAVLYGK----- 307
DQ 588 DQ-----LLOQLTELQSERFQTSQVYTDRELRTRDLQRIQYEGTETELARDVALCAA 639
QY 308 ---RGKYKEA-----EPLQORALE-----IREKVLGTHNPDVAKOLNNLALLCQNOGKYEAV 356
DQ 640 QEMEARHAAVPHLOLLELATEWEDALRERALARERDEAAAEALDAASTSON-ARESAC 698
QY 357 ERYQORALAYEGOLGPDNPVARTKNL-----ASCYLKQCKYAEATLYKEIL-TRAH 410
DQ 699 ER-----LTSLEQQLRESEERAAELASOLEATAAKSSAEQDRESTRATLEQQLRESEAR 753
QY 411 VQFEGSVDDHDKPIWMAHEEREMSKSRHHEGTPYAEYGGYKACKVSSPTVNTTLNL 470
DQ 754 AAELASOLEATAAKSSAEQDRESTRATLEQQLRDSSE-----RAELASOLESTAAKM 808
QY 471 GALLYRROGKLEAAETLEECALRSRQGTDPISO---TKVAELLGESDGRRTS-----Q 520
DQ 809 SA---EQDRESTRATLEQQLRDSSEERAAELASOLESTAAKSSAEQDRESTRATLEQQLR 865
QY 521 EGPDSVKFEGGEDASVAVWMSGSGTQORSGLKIRDVLRRSELLVRKLOGETPRP 580
DQ 866 ESEERAAELASOLESTAAKSSAEQDRESTRATLEQQLRDSSEERAAE-LASOLEATAAK 924
QY 581 SSSNMKRAASLNYLNQ 596
DQ 925 SSAEQDRENTRAALEQ 940

RESULT 15
ID R27204 standard; Protein; 1805 AA.
AC R27204;
DT 15-FEB-1993 (first entry)
DE Rat nestin.
KW Intermediate filament; central nervous system; brain tumour; neurofilament.
OS Rattus norvegicus.
PN W09214821-A.
PD 03-SEP-1992.
PF 21-FEB-1992; U01375.
PR 22-FEB-1991; US-660412.
PA (MASI ) MASSACHUSETTS INST TECHNOLOGY.
PI Lendahl U, McKay RDG;
DR WFI; 92-316175/38.
DR N-PSDB; Q28398.
PT Diagnosis of pre-disposition to brain tumours - using DNA encoding nestin which distinguishes neural multi-potential stem cells from neuronal, glial and muscle cells
PS Claim 3; Fig 1; 63pp; English.
CC The amino acid sequence of the nestin gene which was deduced from the nucleotide sequence, suggests that nestin is a member of the intermediate filament protein family. The rat nestin amino acid sequence shows 75% similarity with the human nestin sequence. There is more than 60% identity between the two sequences. Antibodies to nestin protein can be used in vivo diagnosis of brain tumours.
CC See also Q28399 for the human nestin gene.
SQ Sequence 1805 AA;

Query Match 4.7%; Score 149.5; DB 1; Length 1805;
Best Local Similarity 20.8%; Pred. No. 0.00082;
Matches 165; Conservative 100; Mismatches 230; Indels 239; Gaps 37;

```

QY 6 LGORDEPAGHRLSQE--EILGSTRLVSOGLRLRS-----EQAVLQSLQSOTIECLOQG 57  
Db 951 LEKESQDSGKSLEDESQETFGP--LEKENASLSLAGQOQOEQKLEQETQOQTLRAV---1005  
QY 58 GHEGLVH--EKA-----RQLRRSM--ENIELGLSEAOVMLALASHLST-----97  
Db 1006 GNEQMAVSPPEKVPDPPLPKPLGNDQEIARSLGKNOESLSLKEKGIETVKSLETEIIEP 1065  
QY 98 VESEKOKLRAQVRRLCQENWL-----RDELAGTOORLORSEQAVAOLEEKKH 146  
Db 1066 LETAEEDLERRKSIDTQELMTEVARETVEPPEPPGSLGSVDENRETILTSLEKESOE 1125  
QY 147 LEFLG-----QLRQYDEGHTSEEKEG--DATKDSLDLDFPNEEEEDPSNGLSR 193  
Db 1126 LSSLGKWNVETRVDSQQCLQVEEGLEQHQESLREVKQLPSSGNQOQWEDVVEGKAV 1185  
QY 194 QOGATRA-----QGGYEIPARLTHNLVVIQYAAQGR---YEVAVPLC 234  
Db 1186 GQAPLATTGVGTEDKAELHLRGOGGEEAAAGELLQDIVGEAWSLSGSEPKQRVP--1243  
QY 235 KOALEDLERTSGRHPDVATMLNIALVYRDONKYK-----EAAHLLNDALS 281  
Db 1244 AEALDNLE-----GGALEVPAQSMPEVTEREDRAQAGEQDSIEVTGLGEAA---RTGLE 1296  
QY 282 IRESTLGPDPH-----AVAATLNLAV-----LYGKRGKYKEAPLCQRALEIREK 327  
Db 1297 LEQEVVGLDPRHFAREEAIPPSLGSSESVKAKIAQGLEGGKPKPEKAGALDSGILEL---1353  
QY 328 VLGTNHPDVAKQLNHLALCONQKYEAVERYQRALAI-----YEGQLGPDNPVARTKN 383  
Db 1354 -----PKTSSE-----ALSCQHESESEMEGWESEASLETSDEHGSQDAP--QRPPEPETEE 1402  
QY 384 NLASCYLKQCKYAEATLYK-----EILTRAH-----VQEFSGVDDDHKPIWMHAAE 430  
Db 1403 DEGA---QAALTAPGPKLLEPCSPILTDALHELQPOAEGIOEAGWPQEAQSEALERVEN 1459  
QY 431 REEMSKSRHHEG-----GTPYAEYGGWKACKYS-----459  
Db 1460 EPEFGLGEIPEGLODWEERESEADDLGETLPDSTPL-----GLYLRSAPSKWDLAGEQ 1515  
QY 460 --SPTVNTTLRLNLG--ALYRROGKLEAAETLEECALRSRROCTD-----499  
Db 1516 RLSPQGDAGKEDWGPAPVAAOGLSGPPEEEEE-----QGHGSDLSSEEFEDLGTASLLP 1570  
QY 500 -----PISQTKVAELLGESDG-----RTSQEGPGDSVKFEGGEDASVAVEW 541  
Db 1571 GVPKEVADHVGPVPLQACWDQGESDGFADSEESGEDEEDADEGAESGA---QW 1627  
QY 542 SGDGSGLTQRSGLKIRDVLRSSSELL---VRKLQTEPRPSSNNKRAASLYNLINQ 597  
Db 1628 WSGGA-----SGGCKVQDIAQRGDPVQESVGSGLWDGLRGAANY---PALEMVSD 1679  
QY 598 SAAPLQVSRGLSAS 611  
Db 1680 SAEPGSESEAS 1693

Search completed: August 15, 2000, 10:52:22  
Job time: 25293 sec

Result No.	Score			Query		Length	DB	ID	Description
	Match	Score	Match						
1	167.5	5.3	2101	1	US-08-466-390-4	Sequence 4, Appli			
2	167.5	5.3	2101	1	US-08-470-950-4	Sequence 4, Appli			
3	167.5	5.3	2101	1	US-08-467-781-4	Sequence 4, Appli			
4	167.5	5.3	2101	1	US-08-195-487-4	Sequence 4, Appli			
5	167.5	5.3	2101	2	US-08-483-924-4	Sequence 4, Appli			
6	167.5	5.3	2101	4	PCF-US93-06160-4	Sequence 2, Appli			
7	163	5.2	981	2	US-08-649-046-2	Sequence 4, Appli			
8	160	5.1	885	2	US-08-533-306A-4	Sequence 4, Appli			
9	160	5.1	885	2	US-08-742-923A-4	Sequence 4, Appli			
10	149.5	4.7	955	1	US-08-006-676B-1	Sequence 1, Appli			
11	149.5	4.7	955	1	US-08-282-84-2	Sequence 2, Appli			
12	149.5	4.7	955	2	US-08-428-414A-3	Sequence 3, Appli			
13	149.5	4.7	955	4	PCF-US94-00324-1	Sequence 1, Appli			
14	149.5	4.7	1805	1	US-07-853-913-2	Sequence 2, Appli			
15	148	4.7	2482	1	US-08-328-254-6	Sequence 6, Appli			
16	148	4.7	3248	1	US-08-353-700-1	Sequence 1, Appli			
17	148	4.7	3248	4	PCF-US95-16216-1	Sequence 1, Appli			
18	147	4.7	2154	2	US-08-841-349-4	Sequence 4, Appli			
19	144	4.6	816	2	US-08-533-306A-6	Sequence 6, Appli			
20	144	4.6	816	2	US-08-742-923A-6	Sequence 6, Appli			
21	139	4.4	477	1	US-08-402-217A-3	Sequence 3, Appli			
22	139	4.4	477	1	US-08-700-178-3	Sequence 3, Appli			
23	139	4.4	477	3	US-08-995-654-3	Sequence 3, Appli			
24	137	4.3	546	2	US-08-533-669A-2	Sequence 2, Appli			
25	136	4.3	576	2	US-08-533-306A-2	Sequence 2, Appli			
26	136	4.3	576	2	US-08-742-923A-2	Sequence 2, Appli			
27	135.5	4.3	1354	3	US-07-855-871-2	Sequence 2, Appli			
28	129	4.1	1618	1	US-07-853-913-4	Sequence 4, Appli			





```

Db 1074 AAQIKLELEURQTVKOLKEQLAKKEHAGSGAOSERAGRTPTGPKLEALRAEVSKLE 1133
QY 345 LLCQNGKY-EAVERYIQ-----RALAIYEQ-----LGPDPNPVARTKNNLA 386
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1134 QOCQKQOEQADSLERSLEAERASRAERDSALETLQGLEKAQELGHOSALASAQRELA 1193
QY 387 SCYLKQKGYAEATLYK-----EILTRAHVQEFSGVDDDHK 422
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1194 AFTKQVDSKAEDEKWAQVARGQEAERKNLSLSLEEEVISLNRQVLEKEGESKELKR 1253
QY 423 PIWMAHEEREEMSKSRHHEGGTPYAEYGGWKYKACKVSVPTVNTILRNILGALYRRQKLEA 482
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1254 LVMAESEKSKLEES-----CACCRRQRPATVPPELQNAALLCGRRCRASG 1298
QY 483 AETILECALBSRRQGTDPISQTKVAELLGESDGRRTSQEGPGDSVKFPEGGEDASVAVEMS 542
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1299 REA-EKORVASENLRQELTSQAERAELG-----DELKAWQEKFFQKSOALSTJOLE 1349
QY 543 GDGSGTIL-----ORSGLSGKIR-DVLRRSSEL--L 569
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1350 HTSTQALVSELLPAKHLCOOLQAEAAAEEKHRELEQSKQAAGGLRAELLRAQRELGEL 1409
QY 570 VRKLOGTEPPSSNNKRAASLYNLNQPAAAPLOVSRGLSA 610
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1410 IPLRKQVAEQRTAQOLRAEKASYPAEQLSM--LKKAHGLLA 1448

RESULT 4
US-08-195-487-4
; Sequence 4, Application US/08195487
; Patent No. 5783403
; GENERAL INFORMATION:
; APPLICANT: TOUTATLY, GARY
; APPLICANT: LIDGARD, GRAHAM P
; TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE
; TITLE OF INVENTION: INTERIOR NUCLEAR MATRIX
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TESTA HURWITZ & THIBEAULT
; STREET: 53 STATE STREET
; CITY: BOSTON
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/195,487
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/901,701
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: PITCHER ESQ. EDMUND R
; REGISTRATION NUMBER: 27,829
; REFERENCE/DOCKET NUMBER: MTP-013
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/248-7000
; TELEFAX: 617/248-7100
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2101 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-195-487-4

```



QY 543 GDGSGTL-----QRSGSLGKIR-DVLRSSSEL--L 569  
Db 1350 HTSQALVSELLPAKHLCCQLQAEQAAEKHRELEQSKQAAGGLRAELRAQRELDEL 1409  
QY 570 VRKLGTEPSPSSNNKRAASLNLYNQBPAAPIQVSRGLSA 610  
Db 1410 IPLRQKVAEQRTAQOLRAEKASYAQLSM--LKAHGLLA 1448

RESULT 6  
PCT-US93-06160-4  
Sequence 4: Application PC/TUS9306160  
GENERAL INFORMATION:  
APPLICANT: PITCHER ESO, EDMUND R  
TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE  
NUMBER OF INVENTIONS: 6  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESS: TESTA HURWITZ & THIBEAULT  
STREET: 53 STATE STREET  
CITY: BOSTON  
STATE: MA  
COUNTRY: USA  
ZIP: 02109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: IBM PC compatible  
SOFTWARE: Patent in Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US93/06160  
FILING DATE: 19930621  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: PITCHER ESO, EDMUND R  
REGISTRATION NUMBER: 27,829  
REFERENCE/DOCKET NUMBER: MTP-013  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617/248-7000  
TELEFAX: 617/248-7100  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2101 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
PCT-US93-06160-4

Query Match 5.3%; Score 167.5; DB 4; Length 2101;  
Best Local Similarity 20.5%; Pred. No. 5.9e-06;  
Matches 156; Conservative 106; Mismatches 272; Indels 227; Gaps 32;

QY 15 HRLSOEILGSTRVLSQGLEA-----LRSEHQAVLSLSQTI-----ECLQ--- 55  
Db 750 HKRKEKE-LEERAGRKGLRLQLGAEHAEVLRRLAEAPMAAQAHTAESECEQLVK 808  
QY 56 -----QGGHEE-----GLVHEKARQLRRSMENTELGSLQAQVYMLA-LASH-----LS 96  
Db 809 EVAWRDGYDSQOEAEQYAGMFOELMTLKECEKARQQLQAEKQVAGIESHLSQIS 868  
QY 97 IVESKQKLRQAVRLQOE-----NQWLELAGTQQLRQSEQAQVLE----- 141  
Db 869 RQNKLAELHANLARALQOQVQKEVRAQKJADLSTLQEKMAATSEKVARLETLVKRAKE 928  
QY 142 -BEKKHLEFLGO-LROYDEGHTSEKEGD---ATKDSLDLFPNEEEDPNSGLSRGOG 196  
Db 929 QOETASRELKVPARAGDQFENLEQGGQFCSTQALQAM--ERAEQMGNELELURA 986  
QY 197 ATAAQOQGYEIPARLTHNLVIAAQQHYEVANVPICKQALDLELRTSGRHDPVATWL 256  
Db 987 ALMESQGG-----QOEEERGQOE-----REVARLTQERGRAQADIALEK 1024

QY 257 NILA-LVTRDQNKYE-----AAHLNDALSIRESTLQPDHPAPVAATLNINLAVLYG-K 307  
Db 1025 AARAELEMLQNALNEQORVEFATQELALAHALTEKE---GKD-----QELAKLUGLE 1073  
QY 308 RCKYKEAPLQORALEIREKV-----LGTNHPDYAK-----QLNNLA 344  
Db 1074 AAIKELSELRLQTVKQLEKQELAKKEKEHAGSGAQSAAGRTPTGPKLEALRAEVSKLE 1133  
QY 345 LLCQNGKY-EAVERYQO-----RALAIYEQO-----LGDMPNPNVARTKNLA 386  
Db 1134 QOCQKQEQOQADSLERSLEAERASRAERDSALETLQGLEKAEQELGHSQSALASARELA 1193  
QY 387 SCYLKQGYAEATLYK-----EILTRAHQEFGSVDDHDK 422  
Db 1194 APTKQVQHSKRAEDWKQAQVARGQEAERKNLSLSSLEEVSLNQVLEKEGESKELKR 1253  
QY 423 PIWHAEREEMSKSRHHEGTPYAEGYGGYKACKYSSPTVTNLTNRNLGALYRQCKLEA 482  
Db 1254 LVMAESEKSKLEES-----CACCRQRPATVPQLQNAALLCGRRCRASG 1298  
QY 483 AETLEECALRSRQGTDPISQTKVAELLGSDGRTTSQEGPGDSVKFEGEDASVAVWS 542  
Db 1299 REA-EKQVASENLQELTSQAEARELG-----QELKAWQEFQKEQALSTLQLE 1349  
QY 543 GDGSGTL-----QRSGSLGKIR-DVLRSSSEL--L 569  
Db 1350 HTSQALVSELLPAKHLCCQLQAEQAAEKHRELEQSKQAAGGLRAELRAQRELDEL 1409  
QY 570 VRKLGTEPSPSSNNKRAASLNLYNQBPAAPIQVSRGLSA 610  
Db 1410 IPLRQKVAEQRTAQOLRAEKASYAQLSM--LKAHGLLA 1448

RESULT 7  
US-08-649-046-2  
Sequence 2: Application US/08649046  
Patent No. 5912415  
GENERAL INFORMATION:  
APPLICANT: OLSZEWSKI, NEIL E.  
TITLE OF INVENTION: THE SPINDLY GENE, METHODS OF  
IDENTIFICATION AND USE  
NUMBER OF SEQUENCES: 15  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MUEITING, RAASCH, GEBHARDT & SCHWAPPACH, P.A.  
STREET: 119 NORTH FOURTH STREET, SUITE 203  
CITY: MINNEAPOLIS  
STATE: MINNESOTA  
COUNTRY: USA  
ZIP: 55401  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: IBM PC compatible  
SOFTWARE: Patent in Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/649,046  
FILING DATE: 16-MAY-1996  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: MCCORMACK, MYRA H.  
REGISTRATION NUMBER: 36,602  
REFERENCE/DOCKET NUMBER: 110.00340101  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 612-305-1225  
TELEFAX: 612-305-1228  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 981 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein





Db 588 DQ-----LLOQTELOSERTQLSQVVDTRERLTRLQRIQYEGTETELARDVALCAA 639  
Qy 308 ---RGKYKEA-----EPIQORALE-----IREKVLGTHNPDVAKQLNNLALCQNOQKYEAV 356  
Db 640 QEMEARYHAAVFHLQTLLELATEWEDALRERALEDEAAAAELDAAASTSQN-ARESAC 698  
Qy 357 ERYQORALAIYEGOLGPDNPVARTKNNL-----ASCYLKQKGYAEATLYKEIL-TRAH 410  
Db 699 ER-----LTSLEQOLRESEERAAELASOLEATAAKSSAEQDRENTATLEQOLRESEAR 753  
Qy 411 VQEGSVDDHDKPIWHAEREEMSKSRHHEGGTPYAEYGGWKYACKVSSPTVNTTLNL 470  
Db 754 AAELASOLEATAAKSSAEQDRENTATLEQOLRDSE-----RAELASOLESTTAAKM 808  
Qy 471 GALYRROCKLEAAETLEECALSRROGTDPIQ---TKVAELGESDGRRTS-----Q 520  
Db 809 SA-----EQDRESTRATLEQOLRDSEERAAELASOLESTTAAKSSAEQDRESTRATLEQOLR 865  
Qy 521 EGPDSVKFEGGEDASVAVWSDGSGTLQKRSGLKIRDLVLRSSSELLVRKLOGTEPRP 580  
Db 866 ESEERAAELASOLESTTAAKSSAEQDRESTRATLEQOLRDSEERAAE-LASOLEATAAK 924  
Qy 581 SSSNNKRAASLYNQ 596  
Db 925 SSAEQDRENTAALEQ 940

## RESULT 11

US-08-282-845-2  
; Sequence 2, Application US/08282845  
; Patent No. 5719263  
; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven G.  
; TITLE OF INVENTION: A 230kd Antigen Present in Leishmania  
; NUMBER OF SEQUENCES: 3  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Immunex Corporation  
; STREET: 51 University Street  
; CITY: Seattle  
; STATE: WA  
; COUNTRY: USA  
; ZIP: 98101  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: Apple Macintosh  
; OPERATING SYSTEM: Apple Macintosh Operating System 7.1  
; SOFTWARE: Microsoft Word for Macintosh 5.1a  
; CURRENT APPLICATION DATA:  
; FILING DATE: US/08/282,845  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/006,676  
; FILING DATE: JANUARY 15, 1993  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Perkins, Patricia Anne  
; REGISTRATION NUMBER: 34,693  
; REFERENCE/DOCKET NUMBER: 5004-A  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (206)587-0430  
; TELEFAX: (206)233-0644  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 955 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-282-845-2

Query Match 4.7%; Score 149.5; DB 1; Length 955;

Best Local Similarity 20.9%; Pred. No. 7.1e-05;  
Matches 129; Conservative 92; Mismatches 274; Indels 121; Gaps 25;  
Qy 49 QTIECLOOQGHEGLVH-----EKARLRRSMENIELGLS--EAQVMLASHLS 96  
Db 378 ETLSLTVASRDIVNVAQVNDPPARRIRELEQMEDMQAMAGDPAIVSELKKLA 437  
Qy 97 TVESEKOLRAQVRRRCOE---NQWLDELACTOORLQSRQVAQAEKEKKHLEFLGOL 153  
Db 438 LLESEAKRAADLQALEREHENQVQERLLRATEAEKSELESRAALQE-----EMTATR 492  
Qy 154 ROYDE-DGHTSEKEDGDKSLDDLFNPEEDEDPSNLSRQCGATAAQOGYVPIPARLR 212  
Db 493 ROADKMAALNLRKEQARKER-ELLKEMAKDAA--LSK-----VRRKDAISERE 543  
Qy 213 TLHNLVTOYAAQGR-YEVAVPL-----CKOALEDLERTSGRHPDVATMLNLALVYR 264  
Db 544 KLESTVAQLERQEREREVALDALQTHQKLEALESSERTAAE-----R 587  
Qy 265 DONKYKEAHLNDALSIRE-----STLGPDPHVAATLNLAVLYGK----- 307  
Db 588 DQ-----LLOQTELOSERTQLSQVVDTRERLTRLQRIQYEGTETELARDVALCAA 639  
Qy 308 ---RGKYKEA-----EPIQORALE-----IREKVLGTHNPDVAKQLNNLALCQNOQKYEAV 356  
Db 640 QEMEARYHAAVFHLQTLLELATEWEDALRERALEDEAAAAELDAAASTSQN-ARESAC 698  
Qy 357 ERYQORALAIYEGOLGPDNPVARTKNNL-----ASCYLKQKGYAEATLYKEIL-TRAH 410  
Db 699 ER-----LTSLEQOLRESEERAAELASOLEATAAKSSAEQDRENTATLEQOLRESEAR 753  
Qy 411 VQEGSVDDHDKPIWHAEREEMSKSRHHEGGTPYAEYGGWKYACKVSSPTVNTTLNL 470  
Db 754 AAELASOLEATAAKSSAEQDRENTATLEQOLRDSE-----RAELASOLESTTAAKM 808  
Qy 471 GALYRROCKLEAAETLEECALSRROGTDPIQ---TKVAELGESDGRRTS-----Q 520  
Db 809 SA-----EQDRESTRATLEQOLRDSEERAAELASOLESTTAAKSSAEQDRESTRATLEQOLR 865  
Qy 521 EGPDSVKFEGGEDASVAVWSDGSGTLQKRSGLKIRDLVLRSSSELLVRKLOGTEPRP 580  
Db 866 ESEERAAELASOLESTTAAKSSAEQDRESTRATLEQOLRDSEERAAE-LASOLEATAAK 924  
Qy 581 SSSNNKRAASLYNQ 596  
Db 925 SSAEQDRENTAALEQ 940  
RESULT 12  
US-08-428-414A-3  
; Sequence 3, Application US/08428414A  
; Patent No. 5912166  
; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven G.  
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF  
; TITLE OF INVENTION: LEISHMANIASIS  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SEED AND BERRY  
; STREET: 6300 Columbia Center, 701 Fifth Avenue  
; CITY: Seattle  
; STATE: Washington  
; COUNTRY: USA  
; ZIP: 98104-7092  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/428,414A  
; FILING DATE: 21-APR-1995  
; CLASSIFICATION: 436



Db 809 SA---EQDRESTRATLEQQRDSEBAAELASQLESTTAAKMSAEQDRESTRATLEQQLR 865  
QY 521 EBPQSVFEGEDASVAVESGDSGLQKIRDLVLRSSSELLVKLGQTEPRP 580  
Db 866 ESEERAAELASQLESTTAAKMSAEQDRESTRATLEQQLRDSEBAAE-LASQLEATAAK 924  
QY 581 SSSNMKRAASLYLQ 596  
Db 925 SSAEQDRENTRAALEQ 940  
  
RESULT 14  
US-07-853-913-2  
; Sequence 2, Application US/078533913  
; Patent No. 5338839  
; GENERAL INFORMATION:  
; APPLICANT: McKay, Ronald D.G.  
; TITLE OF INVENTION: Nestin Expression As An Indicator of  
; TITLE OF INVENTION: Neuroepithelial Tumors  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.  
; CITY: Lexington  
; STATE: Massachusetts  
; COUNTRY: U.S.A.  
; ZIP: 02173  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/853,913  
; FILING DATE: 19920319  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/660,412  
; FILING DATE: 22-FEB-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/603,803  
; FILING DATE: 25-OCT-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/201,762  
; FILING DATE: 02-JUN-1988  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/180,548  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Granahan, Patricia  
; REGISTRATION NUMBER: 32,227  
; REFERENCE/DOCKET NUMBER: MIT-4641AAAA  
; TELEPHONE: 617-861-6240  
; TELEFAX: 617-861-9540  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1805 amino acids  
; TYPE: AMINO ACID  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-07-853-913-2  
  
Query Match 4.7%; Score 149.5; DB 1; Length 1805;  
Best Local Similarity 20.8%; Pred. No. 0.0002;  
Matches 165; Conservative 100; Mismatches 290; Indels 239; Gaps 37;  
  
QY 6 LGORDEPAGHRLSQE--EILGSTRSLVSGGLEALRS-----EHOAVQLSQTIECLQOG 57  
Db 951 LEKESQDSGKSLDESQETGCP--LEKENAESLRSAGQOEKLEQETQOTLRAV--- 1005  
QY 58 GHESGLVH--EKA-----RQLRRSM--ENIELGLSEAVQVMLASHLST----- 97

Db 1006 GNEQMAVSPPEYDPELPKPLNGDQEIARSLGKENGESLVSLEKGIETVKSLETEIIEP 1065  
QY 98 VESEKGLRAQYRRLCOENWL-----RDELAGTQORLQSRSEQAQAQLEEKKH 146  
Db 1066 LETAEEDLERRKSIDTQPLWSTEVARTVEPPEDEPGSLGSDVENRETITSLEKESQE 1125  
QY 147 LEFLG-----QLRQYDEGHTSEEKQ--DATKDSLDDLPFNEEEDPSNGLSR 193  
Db 1126 LSSLGNVETRVEDSQOCLQVEEGLOEHOQESLREVKQELPSSGNQQRWEDVVEGRAV 1185  
QY 194 GOGATAA-----QGGYEIPARLTLNHLVLYQAAQGR---YEVAVPLC 234  
Db 1186 GQAPLATTGVTGDKAELHLRGQGEAEAAAGELLQDIVGEAWSLSSESPEKQORVP-- 1243  
QY 235 KOALEDLERTSGRHPDVATMLNIALVYRDQNKY-----EAAHLNDALS 281  
Db 1244 AEAIDNLE-----GGALEVPVAQSMFEVTERDEDRQAQGEQDSIEVTLGLEAA---RTGLE 1296  
QY 282 IRESTGLPDHP-----AVAATLNNLAV-----LYGKRKYKEAEPLCORALEIREK 327  
Db 1297 LEQEVVGLDPRHFAREEAIPPSLGEESVKAKIAQGLEPGKEPKEAGALDSGILEL--- 1353  
QY 328 VLGTHPDYAKQLNNLALICQNGKYEAERYVYQALAI---YEGQLGPDNPVARTKN 383  
Db 1354 -----PKTSSE-----ALECQGHSESEMEGWEEEAASLETSDEHSGSDAP--QRPPETEE 1402  
QY 384 NLASCYLKQKGYAEATLYK-----EILTRAH-----VOEFGSVDDDHKPIWMHAE 430  
Db 1403 DEGA---QAALTAPGPKLLEPCSPILITDAHELQPAQSGIOEAGWQPEAGSEALERVEN 1459  
QY 431 REEMSKSRHHEG-----GTPYAEYGGWYKACKVS----- 459  
Db 1460 EPEFGLGEIPEGLQDWEEGRESEADDLGTLTLPDSTPL-----GLYLRSPASPKWDLAGEQ 1515  
QY 460 --SPTVNTTLRNLG--ALYRROGKLEAAETLEECALRSRQGTD----- 499  
Db 1516 RLSPQGDAGKEDWGPVAPAAQGLSGPPEEEE-----QHGSDLSSEEFEDLGTASLLP 1570  
QY 500 -----PISQTKVAELLGSDG-----RRTSQEGPDSYKFFEGGEDASVAVEM 541  
Db 1571 GVPKEVADHVGVPPVLPQACWDQGGESDGFADDEESGEEDADEGAESGA---QM 1627  
QY 542 SGDSGTLQSRGSLGKIRDLVLRSSSELL-----VRKLGTEPRPSSSNMKRAASLYLQNP 597  
Db 1628 WSGA-----SGGCKYQDIAQRGDPVQESVGVSGLWDDGLRGAANV---PALEMVSOD 1679  
QY 598 SAAPLQVSRGLSAS 611  
Db 1680 SAEPGSGSESESAS 1693  
  
RESULT 15  
US-08-328-254-6  
; Sequence 6, Application US/08328254  
; Patent No. 5710022  
; GENERAL INFORMATION:  
; APPLICANT: Zhu, Xuelliang  
; APPLICANT: Lee, Wen-Hwa  
; TITLE OF INVENTION: A No. 5710022e1 Nuclear Mitotic Phosphoprotein  
; NUMBER OF SEQUENCES: 8  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Campbell and Flores  
; STREET: 4370 La Jolla Village Drive, Suite 700  
; CITY: San Diego  
; STATE: California  
; COUNTRY: USA  
; ZIP: 92122  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25



Search completed: August 15, 2000, 10:53:34  
Job time: 23119 sec

```

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/328,254
; FILING DATE: 24-OCT-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/141,239
; FILING DATE: 22-OCT-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-CJ 1191
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2482 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-328-254-6

Query Match      4.7%; Score 148; DB 1; Length 2482;
Best Local Similarity 21.2%; Pred. No. 0.00047;
Matches 137; Conservative 112; Mismatches 256; Indels 140; Gaps 30;

QY 16 RLS--QEELGSTRVLSQGLALRSQAVLQSLSTIECLQOGHGEGLVHEKA----R 69
Db 1474 RLSSTQEEV---HQLRRGIEKLRVRIEADKKQLHIAEKLRERENDSLDKKVENLER 1529

QY 70 QLRRSMNIELGL-----SEAQV-----MLALASHLSTVESEKQKLRQAVRR 111
Db 1530 ELQMSSENGELVILDAENSKAEVETLKTQTEEMARSLKVFELDLVTLRSEKENLTQI-- 1587

QY 112 LQENQWLRE----LAGTQRLQREQAVALLEKK-HLEFL-GQLRQYDE--DGHTS 163
Db 1588 --QEKOGQISELDKLSFKSLLEKEQAEIQIKESKTAVEMQLQNLKELNEAVALCG 1645

QY 164 EKEGDATKSDLDLPNNEEDPSNGLSRGOGATAAQGGYEIPARLRLHN---LVIQ 220
Db 1646 DQEIWKATEQSLDP--PIEEHQLRNSIEK-----LRARLEADEKKQLCVLQ 1690

QY 221 YAAQGRYEVAVPCKQALDELDER--TSGRGHPDVATMLNIALVYRDQNKYFAAHLND 278
Db 1691 QLKSEHH--ADLLKGRVENLERELEIARTINQEAAL-----EAENSKGE 1733

QY 279 ALSIRESTLGPDPHPAATAINNLAIVLYGKRGYKAEPLCQRALEIREKVLCTNHDPDAK 338
Db 1734 VETLKAKIEGTMQSLRGLELDVVVTIRSEKENTNELQKQERISEL--EIINSSFFENILQ 1791

QY 339 QLNNLALLCQNGKYQAVERYQYRALAIYEGQLGDPNPVARTKNNLASCYLK-QGKYAE 397
Db 1792 E-----KEQEKVQMKK--SSTAMEMLTQLKELNERNVAAHNDQEAQKAEQNLSSQ 1842

QY 398 AETLYKE-----ILTRAHV-----QFSGSVDDDHKPIWMHAER 431
Db 1843 VECLELEKQAQLLQGLDEAKNNIVLOS SVNGLTQEVEDGKQLEKKDEEISRLKNQIQDQ 1902

QY 432 EEM-SKSRHHEGTPYAEYGGWYKACKVSSPTVNTTLNGLALYRR-----QKLEAAET 485
Db 1903 EOLVSKLSQVEG-----EHLNKEQ-----NLELRNLVELEQKIQVLOSKNASLOD 1949

QY 486 LEECALRSRROGTDPISQKVAEL-LGESDGRRTSOEGPGDSVKFEGGEDASVAVEWSDG 544
Db 1950 TLEVLOSSYKNLENELELTMDKMSFVKVNMKTAKE---TELQREHMAQKTAELQEE 2006

QY 545 GSGTLQR-SGSLGKIRDLVRRSELVLR-KLOGTEPRPSSNNMKR 587
Db 2007 LSQEKNLRLAGQLLLEETKSKDKQLKELTLENSELKSLDCMHK 2051
```

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 15, 2000, 04:31:13 ; Search time 131.1 Seconds  
(without alignments)  
292.194 Million cell updates/sec

Title: US-09-036-614A-1  
Perfect score: 3161  
Sequence: 1 MSGVLGQRDEPAGHRLSQE.....APLQVSRGLSASTMDLSSSS 619

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 178050 seqs, 61884766 residues

Total number of hits satisfying chosen parameters: 178050

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_64: \*  
1: pir1: \*  
2: pir2: \*  
3: pir3: \*  
4: pir4: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	2000	63.3	569	I3013	kinesin light chain
2	1985.5	62.8	560	C41539	kinesin light chain
3	1956	61.9	551	B41539	kinesin light chain
4	1954.5	61.8	542	A41539	kinesin light chain
5	1822	57.6	686	S33815	kinesin light chain
6	1821	57.6	686	S36727	kinesin light chain
7	1809.5	57.2	649	S33813	kinesin light chain
8	1803.5	57.1	677	S33814	kinesin light chain
9	1520	48.1	537	T34113	kinesin light chain
10	1517	48.0	563	S47997	kinesin light chain
11	1516	48.0	540	S41864	hypothetical prote
12	1491.5	47.2	522	S41865	kinesin light chain
13	1453	46.0	451	S33816	kinesin light chain
14	965	30.5	516	T23827	hypothetical prote
15	241.5	7.6	114	T37998	kinesin light chain
16	232	7.3	609	T01892	hypothetical prote
17	229.5	7.3	886	T35469	probable ATP /GTP
18	189	6.0	342	E70463	conserved hypothe
19	189	6.0	948	S75931	hypothetical prote
20	185	5.9	1791	T02909	hypothetical prote
21	181	5.7	1939	T18372	repeat organellar
22	179.5	5.7	2176	T13806	teoucan gene protei
23	179	5.7	1963	T20770	hypothetical prote
24	179	5.7	1966	1 MWKV	myosin heavy chain
25	175.5	5.6	671	A40692	signal recognition
26	173.5	5.5	848	A44972	paramyosin - nemat
27	167.5	5.3	2101	A42184	nuclear mitotic ap
28	167.5	5.3	3187	JC5837	364K Golgi complex
29	166.5	5.3	879	A48575	paramyosin - nemat

smooth muscle myos  
myosin heavy chain  
smooth muscle myos  
ninein - mouse  
ro-3 protein - Neu  
myosin heavy chain  
hypothetical prote  
centrosome associa  
hypothetical prote  
hypothetical prote  
pericentrin - mous  
transformation-sen  
hypothetical prote  
purine NTPase homo  
giantin - human  
trfa protein - sll

30 166 5.3 1938 2 JC5421  
31 166 5.3 1972 1 A41604  
32 166 5.3 1972 2 JC5420  
33 163 5.2 2168 2 T30171  
34 162.5 5.1 1300 2 T18364  
35 162 5.1 1969 1 S02771  
36 162 5.1 1992 2 T23622  
37 162 5.1 2442 2 T08621  
38 161.5 5.1 1133 2 T22976  
39 160.5 5.1 1426 2 T00337  
40 160 5.1 1920 2 A3188  
41 159.5 5.0 543 2 A38093  
42 159 5.0 1974 2 T30010  
43 157 5.0 886 2 H69378  
44 157 5.0 3225 2 I52300  
45 156.5 5.0 1390 2 T14004

ALIGNMENTS

RESULT 1

I53013  
kinesin light chain - human

C:Species: Homo sapiens (man)

C:Date: 02-Jul-1996-#sequence\_revision 02-Jul-1996 #text\_change\_05-Nov-1999

C:Accession: I53013

R:Cabeza-Arvelaiz, Y.; Shih, L.C.; Hardman, N.; Asselbergs, F.; Bilbe, G.; Schmitz, A

DNA Cell Biol. 12, 881-892, 1993

A:Title: Cloning and genetic characterization of the human kinesin light-chain (KLC)

A:Reference number: I53013; MUID:94099888

A:Accession: I53013

A:Status: preliminary; translated from GB/EMBL/DBDJ

A:Molecule type: mRNA

A:Residues: 1569 <RES>

A:Cross-references: GB:I04733; NID:G307084; PID:AAA16576.1; PID:G307085

Query Match 63.3%; Score 2000; DB 2; Length 569;

Best Local Similarity 69.8%; Pred. No. 9.1e-109;

Matches 395; Conservative 77; Mismatches 76; Indels 18; Gaps 8;

QY 1 MSGVLGQRDEPAGHRLSQEILGSTRVLSQGLSEARSEHQAVLQSLSTIECLOQGGH 60  
Db 1 MSTWVIKEDKL--EKLTDIELISKTKQVIOGLEALKNEHNSILOSLLETLCCLKK--DDE 57  
QY 61 EGLVHEKAROLRSMENIELGLSEAQVMLALASHLTVESEKOKLRAQVRRRCQENQWL 120  
Db 58 SNLVEEKSNNIRKSLMLELGLSEAQVMMALSNHNAVESEKOKLRAQVRRRCQENQWL 117  
QY 121 DELAGTQORLQSEQAVALEEKKHLEFLGOLROYDEDEGHTSEEKGDATKDSLDDLEP 180  
Db 118 DELANTQOKLQSEQVLALEEKKHLEFLGOLROYDEDEGHTSEEKGDATKDSLDDLEP 177  
QY 181 NEEEDPSNGLSR--CGGATAAQGGYEIPARLTLNLVIOYAAQGRYEVAVPLCKQAL 238  
Db 178 N-DEDDPGGICIQOQHSSAAAAAQGGYEIPARLTLNLVIOYAAQGRYEVAVPLCKQAL 236  
QY 239 EDLERTSGRHDPDPAVMTNLALVYRDQNKYEAAHLNDALSIRESTLGPDPHVAATL 298  
Db 237 EDLEKTSGDHPDPAVMTNLALVYRDQNKYKDAANLNDALAIREKTLGDKHPAATL 296  
QY 299 NNLAVLYGKRGYKEAPELCCORALEIREKVLGNHDPVAKLNNLALCONQKYEAEV 358  
Db 297 NNLAVLYGKRGYKEAPELCCORALEIREKVLGNHDPVAKLNNLALCONQKYEAEV 356  
QY 359 YTORALAIYEGOLGPDNPVARTKNNLASCYLKQGYAEATLYKEILTRAHVQFSGVD 418  
Db 357 YTORALEIYQTLGPDNPVARTKNNLASCYLKQGYAEATLYKEILTRAHVQFSGVD 416  
QY 419 DQHKPTWMAEEREESKSRHHEGGTPYAEYGGWKACKVSSPTVNTTLRNIGALYRQ 478  
Db 417 DENKPTWMAEEREESKSRHHEGGTPYAEYGGWKACKVSSPTVNTTLRNIGALYRQ 474

### RESULT 3

A;Accession: A41539  
A;Status: preliminary



```
Db 121 WLRLDLAATQKLTSEQNLADELVKYKHLEYMNSIKKYDEDRTPDEEA---SSSDPLDL 177
QY 178 LFPNEEEDPSNGLS-----RQGATAAQGGYEIPARLTLHNLVIQYAAQGRYEV 229
Db 178 GFP-----EDDDGQADSESYPPQOTGSGSVSAAAGGYEIPARLTLHNLVIQYASQRYEV 233
QY 230 AVPLCKQALDELRTSGRHPDVATMLNIALVYRQNKYKEAAHLLNDALSIRESTLGP 289
Db 234 AVPLCKQALDELRTSGRHPDVATMLNIALVYRQNKYKEAGNLLHDALAIKREXTLGP 293
QY 290 DHPAATAATLNNAVLYGKRGKYKEAEPKCRALREIREKVLGKDPDVAKOLNNLALCON 349
Db 294 DHPAATAATLNNAVLYGKRGKYKEAEPKCRALREIREKVLGKDPDVAKOLNNLALCON 353
QY 350 QGKYEAERYQORALAYEGOLGPDNPVARTKNNLASCYLKQGYAEATLYKEILTRA 409
Db 354 QGKYEEVEWYQORALEIYEKLGPDNPVARTKNNLAAAYLKQGYKAAETLYKQVLTTRA 413
QY 410 HVQEFQ-SVDD-DHKPTIWMHAEEREEMSKSRHHEGGTPYAEYGGWKYKACKV-----SSPT 462
Db 414 HERFGLSADDDKKNKPIWQAEEREKGFKN---APYGDYGGWHKAAKVDSSRSPT 470
QY 463 VNTTLNMGALYRQGLKLEAAETLEECALRSRROGTDPISQTKVAELG-----511
Db 471 VTTTLKNGALYRQGYDAAEILEECAMKSRNALDMVRETKVRELLGQDLSTDVPRSE 530
QY 512 -----ESDGRRTS---QEGPGDSVKFE---GGEDASVAVESG---SPT 543
Db 531 AMAKERHRRSSGTPRHSSTESVYEKTDGSEE-----NGDGKLRSGSLKURAS 581
QY 544 -----DGSGLQSRGSLKIRDLVRRSELLVRKLOGTGTEPPRSSNMKRAASLNY 593
Db 591 PRSPHVLVNGDKLRRSGSLKIRASVRRSSTKLLNKLKGRE-SDDGGMKRASSMSV 649
QY 594 L-----NOPSAAAPLOVS-RGLSASTMDLSS 617
Db 650 LPSRGNDSTPAPILQSRGRVGSNDLSS 679

RESULT 7
S33813
kinesin light chain - sea urchin (Strongylocentrotus purpuratus)
C:Species: Strongylocentrotus purpuratus (purple urchin)
C:Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 09-Sep-1997
C:Accession: S33813
J.Wedman, K.P.; Knight, A.E.; Kendrick-Jones, J.; Scholey, J.M.
A:Title: Sequences of sea urchin kinesin light chain isoforms.
A:Reference number: S33813; MUID:93267648
A:Accession: S33813
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-649 <WED>
A:Cross-references: EMBL:L10233; NID:g161525; PID:g161526

Query Match 57.2%; Score 1809.5; DB 2; Length 649;
Best Local Similarity 59.2%; Pred. No. 1.2e-97;
Matches 392; Conservative 81; Mismatches 124; Indels 65; Gaps 16;

QY 1 MSGVLGQRDEPAGHR--LSQEEILGSTRVLSQGLEALRSEHQAVLSQSLTIECLQGG- 57
Db 1 MSGSKLSTPNNSGGGGLSQEQIITGTREVTKGLEQKNEHNDILNSLYQSLKMLKDD 60
QY 58 GHEEGLVHEKARQLRSMENIELGSEAQVMLALASHLSTVESEKQKLRQAVRRLCOENQ 117
Db 61 PGDSNLVVEKTDIIKESLESLELGGEAKVMALGHLLNMVYAEKQKLRQAVRRLVQENT 120
QY 118 WLRLDLAATQKLTSEQNLADELVKYKHLEYMNSIKKYDEDRTPDEEA---SSSDPLDL 177
Db 121 WLRLDLAATQKLTSEQNLADELVKYKHLEYMNSIKKYDEDRTPDEEA---SSSDPLDL 177
QY 178 LFPNEEEDPSNGLS-----RQGATAAQGGYEIPARLTLHNLVIQYAAQGRYEV 229
Db 178 LFPNEEEDPSNGLS-----RQGATAAQGGYEIPARLTLHNLVIQYAAQGRYEV 229
```

```
Db 178 GFP-----EDDDGQADSESYPPQOTGSGSVSAAAGGYEIPARLTLHNLVIQYASQRYEV 233
QY 230 AVPLCKQALDELRTSGRHPDVATMLNIALVYRQNKYKEAAHLLNDALSIRESTLGP 289
Db 234 AVPLCKQALDELRTSGRHPDVATMLNIALVYRQNKYKEAGNLLHDALAIKREXTLGP 293
QY 290 DHPAATAATLNNAVLYGKRGKYKEAEPKCRALREIREKVLGKDPDVAKOLNNLALCON 349
Db 294 DHPAATAATLNNAVLYGKRGKYKEAEPKCRALREIREKVLGKDPDVAKOLNNLALCON 353
QY 350 QGKYEAERYQORALAYEGOLGPDNPVARTKNNLASCYLKQGYAEATLYKEILTRA 409
Db 354 QGKYEEVEWYQORALEIYEKLGPDNPVARTKNNLAAAYLKQGYKAAETLYKQVLTTRA 413
QY 410 HVQEFQ-SVDD-DHKPTIWMHAEEREEMSKSRHHEGGTPYAEYGGWKYKACKV-----SSPT 462
Db 414 HERFGLSADDDKKNKPIWQAEEREKGFKN---APYGDYGGWHKAAKVDSSRSPT 470
QY 463 VNTTLNMGALYRQGLKLEAAETLEECALRSRROGTDPISQTKVAELG-----511
Db 471 VTTTLKNGALYRQGYDAAEILEECAMKSRNALDMVRETKVRELLGQDLSTDVPRSE 530
QY 512 -----ESDGRRTS---QEGPGDSVKFE---GGEDASVAVESG---SPT 543
Db 531 AMAKERHRRSSGTPRHSSTESVYEKTDGSEE-----NGDGKLRSGSLKURAS 581
QY 544 -----DGSGLQSRGSLKIRDLVRRSELLVRKLOGTGTEPPRSSNMKRAASLNY 593
Db 591 PRSPHVLVNGDKLRRSGSLKIRASVRRSSTKLLNKLKGRE-SDDGGMKRASSMSV 649
QY 594 L-----NOPSAAAPLOVS-RGLSASTMDLSS 617
Db 650 LPSRGNDSTPAPILQSRGRVGSNDLSS 679

RESULT 8
S33814
kinesin light chain - sea urchin (Strongylocentrotus purpuratus)
C:Species: Strongylocentrotus purpuratus (purple urchin)
C:Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 09-Sep-1997
C:Accession: S33814
J.Wedman, K.P.; Knight, A.E.; Kendrick-Jones, J.; Scholey, J.M.
A:Title: Sequences of sea urchin kinesin light chain isoforms.
A:Reference number: S33813; MUID:93267648
A:Accession: S33814
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-677 <WED>
A:Cross-references: EMBL:L10234; NID:g161527; PID:g161528

Query Match 57.1%; Score 1803.5; DB 2; Length 677;
Best Local Similarity 57.9%; Pred. No. 2.8e-97;
Matches 394; Conservative 82; Mismatches 130; Indels 75; Gaps 16;

QY 1 MSGVLGQRDEPAGHR--LSQEEILGSTRVLSQGLEALRSEHQAVLSQSLTIECLQGG- 57
Db 1 MSGSKLSTPNNSGGGGLSQEQIITGTREVTKGLEQKNEHNDILNSLYQSLKMLKDD 60
QY 58 GHEEGLVHEKARQLRSMENIELGSEAQVMLALASHLSTVESEKQKLRQAVRRLCOENQ 117
Db 61 PGDSNLVVEKTDIIKESLESLELGGEAKVMALGHLLNMVYAEKQKLRQAVRRLVQENT 120
QY 118 WLRLDLAATQKLTSEQNLADELVKYKHLEYMNSIKKYDEDRTPDEEA---SSSDPLDL 177
Db 121 WLRLDLAATQKLTSEQNLADELVKYKHLEYMNSIKKYDEDRTPDEEA---SSSDPLDL 177
QY 178 LFPNEEEDPSNGLS-----RQGATAAQGGYEIPARLTLHNLVIQYAAQGRYEV 229
Db 178 LFPNEEEDPSNGLS-----RQGATAAQGGYEIPARLTLHNLVIQYAAQGRYEV 229
```

QY 230 AVPLCKQALDERTSGRSHGPDVATMLNIALVTRDQNKYKEAAHLNDALSTRESTLGP 289  
 DB 234 AVPLCKQALDERTSGRSHGPDVATMLNIALVTRDQNKYKEAGNLDHALIREKTLP 293  
 QY 290 DHPAATLNNLAVLYGKRGYKEAPLQCORALEIREKVLGTNHPDPAKQNLNALLCQN 349  
 DB 294 DHPAATLNNLAVLYGKRGYKEAPLQCORALEIREKVLGTNHPDPAKQNLNALLCQN 353  
 QY 350 QKYEAVERYQALAIYEGQLGPDNPVARTKNNLASCYLKQKYAEATLYKEILTRA 409  
 DB 354 QKYEAVERYQALAIYEGQLGPDNPVARTKNNLASCYLKQKYAEATLYKEILTRA 413  
 QY 410 HYQEFQ-SVDD-DHKPTWMAEEREMSKSRHHEGGTPYAEYGGWKYKACKY-SSPT 462  
 DB 414 HEREFGLSADDKDNFPIWMAEEREMSKSRHHEGGTPYAEYGGWKYKACKY-SSPT 470  
 QY 463 VNTTLNGLYRROCKLEAATLBEALRRSROGTDPISTKVAELLG----- 511  
 DB 471 VNTTLNGLYRROCKLEAATLBEALRRSROGTDPISTKVAELLG----- 530  
 QY 512 ---ESDGRRTS---QEGPDSVKFE---GGEDASVAVES----- 542  
 DB 531 ANAKERHRRSGTFRHSGTESVSEKTDGSEERKAKDRSRSPAGYVEIPRSPHVLV 590  
 QY 543 GDSGLTQSRGSLGKIRDLRRSSBELLVRLKQTEPRPSSNMKRAASLNYL-----NQP 597  
 DB 591 ENGDKLRRSGSLKLRASVRSSTKLLKLGRE-SDDGGMKRASSMSVLPGRGNDES 649  
 QY 598 SAAPLOVS-RGLSASTMDLSS 617  
 DB 650 TPAPIQLSGRVRGSHDNLSS 670

RESULT 9  
 T34113  
 hypothetical protein C18C4.10 - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C:Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 29-Oct-1999  
 C:Accession: T34113  
 R:Gatting, S.  
 submitted to the EMBL Data Library, April 1996  
 A:Description: The sequence of C. elegans cosmid C18C4.  
 A:Reference number: 221478  
 A:Accession: T34113  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-537 <GAT>  
 A:Cross-references: EMBL:U55369; PIDN:AAC25830.1; GSPDB:GN00023; CESP:C18C4.10  
 A:Experimental source: strain Bristol N2; clone C18C4  
 C:Genetics:  
 A:Gene: CESP:C18C4.10  
 A:Map position: 5  
 A:Introns: 37/1; 142/3; 449/2; 485/3

Query Match 48.1%; Score 1520; DB 2; Length 537;  
 Best Local Similarity 58.4%; Pred. No. 5, 7e-81;  
 Matches 313; Conservative 75; Mismatches 106; Indels 42; Gaps 9;

QY 17 LSQEEILGSTRIVSQGLEALRSEHQAVLSQTIETCLOQGGHEGLVHEKARLRSME 76  
 DB 1 MSQDDVTTLGRTVQOQGLEALRSEHSTISNTLETSVKGYKE--DEAPLPKQLSQINDNLD 58  
 QY 77 NIELGLSEAQVLMALASHLSTVESEKOKLRAQVRRRCQENQWLRLDELQAGTOQLRQSEA 136  
 DB 59 KIVCGVDETSMLMVFQLTQGDMAHQYQARRRRCQENAWLRDELSTQIKLQSQEQM 118  
 QY 137 VAQLEEKHLEFLGQLRQYDEGHTSEEKQ-----DATKSLDDL-FPNEEED--- 186  
 DB 119 VAQLEENKHLKYMASIKQFD-DGTQSDTKTSVDVGPQPVNTLQELGFGPEDEDMNA 177  
 QY 187 -----PSNGLSRGOGATAAQGGYEIPARLTLNVLVIOYAAQGRYEVAVPLCKQALE 239

DB 178 SQFNQPTPANOMA-----ASANVGYEIPARLTLNVLVIOYAAQGRYEVAVPLCKQALE 231  
 QY 240 DLERTSGRSHGPDVATMLNIALVTRDQNKYKEAAHLNDALSTRESTLGPDPHVAATLN 299  
 DB 232 DLEKTSGRHDPVATMLNIALVTRDQNKYKEAANLNALSIREKCLGESHHPAATLN 291  
 QY 300 NLAVLYGKRGYKEAPLQCORALEIREKVLGTNHPDPAKQNLNALLCQNGKYEAVERY 359  
 DB 292 NLAVLYGKRGYKEAPLQCORALEIREKVLGTNHPDPAKQNLNALLCQNGKYEAVERY 351  
 QY 360 YQALAIYEGQLGPDNPVARTKNNLASCYLKQKYAEATLYKEILTRAHQVEFGSVD 419  
 DB 352 YKRALEIYESKLGPDNPVARTKNNLSSAYLKQKYKEAEELYKQILTRAHEREFGQISG 411  
 QY 420 DHKPTWMAEEREMSKSRHHEGGTPYAEYGGWKYKACKYSSPTVNTLNLGALYRROCK 479  
 DB 412 ENKPTWMAEEREMSKSRHHEGGTPYAEYGGWKYKACKYSSPTVNTLNLGALYRROCK 468  
 QY 480 LEAAETLEECALRRQ-----GTDPISTKVAELLGSDG---RRTSQEG 522  
 DB 469 YEAATLEEDVALRAKQHEPLRSGAMGIDEMSQSMASITIGGSRNSMTTSTQSG 524

RESULT 10  
 S47997  
 kinesin light chain (isoform 1) - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C:Date: 07-May-1995 #sequence\_revision 21-Jul-1995 #text\_change 21-Jul-1995  
 C:Accession: S47997  
 R:Fah, J.; Amos, L.A.  
 J. Mol. Biol. 240, 507-512, 1994  
 A:Title: Kinesin light chain isoforms in Caenorhabditis elegans.  
 A:Reference number: S47997; MUID:94322393  
 A:Accession: S47997  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-563 <PAN>  
 A:Cross-references: EMBL:Z29644

Query Match 48.0%; Score 1517; DB 2; Length 563;  
 Best Local Similarity 56.7%; Pred. No. 9, 1e-81;  
 Matches 317; Conservative 75; Mismatches 125; Indels 42; Gaps 9;

QY 17 LSQEEILGSTRIVSQGLEALRSEHQAVLSQTIETCLOQGGHEGLVHEKARLRSME 76  
 DB 4 MSQDDVTTLGRTVQOQGLEALRSEHSTISNTLETSVKGYKE--DEAPLPKQLSQINDNLD 61  
 QY 77 NIELGLSEAQVLMALASHLSTVESEKOKLRAQVRRRCQENQWLRLDELQAGTOQLRQSEA 136  
 DB 62 KIVCGVDETSMLMVFQLTQGDMAHQYQARRRRCQENAWLRDELSTQIKLQSQEQM 121  
 QY 137 VAQLEEKHLEFLGQLRQYDEGHTSEEKQ-----DATKSLDDL-FPNEEED--- 186  
 DB 122 VAQLEENKHLKYMASIKQLD-DGTQSDTKTSVDVGPQPVNTLQELGFGPEDEDMNA 180  
 QY 187 -----PSNGLSRGOGATAAQGGYEIPARLTLNVLVIOYAAQGRYEVAVPLCKQALE 239  
 DB 181 SQFNQPTPANOMA-----ASANVGYEIPARLTLNVLVIOYAAQGRYEVAVPLCKQALE 234  
 QY 240 DLERTSGRSHGPDVATMLNIALVTRDQNKYKEAAHLNDALSTRESTLGPDPHVAATLN 299  
 DB 235 DLEKTSGRHDPVATMLNIALVTRDQNKYKEAANLNALSIREKCLGESHHPAATLN 294  
 QY 300 NLAVLYGKRGYKEAPLQCORALEIREKVLGTNHPDPAKQNLNALLCQNGKYEAVERY 359  
 DB 295 NLAVLYGKRGYKEAPLQCORALEIREKVLGTNHPDPAKQNLNALLCQNGKYEAVERY 354  
 QY 360 YQALAIYEGQLGPDNPVARTKNNLASCYLKQKYAEATLYKEILTRAHQVEFGSVD 419  
 DB 355 YKRALEIYESKLGPDNPVARTKNNLSSAYLKQKYKEAEELYKQILTRAHEREFGQISG 414  
 QY 420 DHKPTWMAEEREMSKSRHHEGGTPYAEYGGWKYKACKYSSPTVNTLNLGALYRROCK 479

Db 415 ENKPIWQIAEERE---NKHKGEGATANEQAGWAKAAKVDSPVTVTTLKLNGLALYRROGK 471  
Qy 480 LEAETLEECALRSRQ-----GTDPISTQTKVAELLEGSDG---RTSQEGGDS 526  
Db 472 YEAATLEEDVALRAKKQHEPLRSGAMGIDEMSQSMWASTIGGSRNSTTSTQMSQSM 531  
Qy 527 VKFEGGEDASVAVWEGSDG 545  
Db 532 ASTIGGRNSTTSTQ 550

RESULT 11  
S41864  
kinesin light chain (isoform 1) - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 06-Jan-1995 #sequence\_revision 06-Jan-1995 #text\_change 09-Sep-1997  
C:Accession: S41864  
R:Fan, J.; Amos, L.A.  
submitted to the EMBL Data Library, January 1994  
A:Description: Kinesin light chain isoforms in C. elegans.  
A:Reference number: S41864  
A:Accession: S41864  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-540 <FAN>  
A:Cross-references: EMBL:229644; NID:g452601; PID:g452602

Query Match 48.0%; Score 1516; DB 2; Length 540;  
Best Local Similarity 58.4%; Pred. No. 9.9e-81;  
Matches 313; Conservative 74; Mismatches 107; Indels 42; Gaps 9;

Qy 17 LSOEILGSTRVLSQGLEALRSEHQVLSQSLQTECLOQGGHEGLVHEKARQLRRSME 76  
Db 4 MSQDDVTGLRTVQOGLLEALREHSTISNTLETSVKGVKE--DEAPLPKQLSINDNLD 61  
Qy 77 NIELGLSEAQVMLALASHLSTVESEKQKRAQVRLRCQENQWLRLDELQAGTQORLQSEQA 136  
Db 62 KLVCGVDETSMLMVFLQTLQGMDAQHQYQAQRRRLCQENAWLRDELSSSTQIKLQOSEQM 121  
Qy 137 VAOLEEKKHLEFLGQLROYDEDEHTSEEKGG-----DATKDSLDDL-FPNEEED--- 186  
Db 122 VAOLEEENKHLKYMASIKQLD-DGTQSDTKTSVDVGPQVPTNETLQELGFGPEDEEDMNA 180  
Qy 187 -----PSNGLSRGOGATAAQGGYEIPARLTLNLVIOYAAQGYEYAVPLCKQALE 239  
Db 181 SOFNOPTPANQMA-----ASANVGEIPARLTLNLVIOYASQGYEYAVPLCKQALE 234  
Qy 240 DLERTSGRHPDVPATMLNIALVYRDQNKYKEAAHLNDALSTRESTLGPDPHPAVALTN 299  
Db 235 DLEKTSQGHDPDVPATMLNIALVYRDQNKYKEAAHLNDALSTRESTLGPDPHPAVALTN 294  
Qy 300 NLAVLYGRGKYKEAEPICQRALEIREKVLGTHNPDVAKQLNLLALCQNOGKYEAVERY 359  
Db 295 NLAVLYGRGKYKEAEPICQRALEIREKVLGTHNPDVAKQLNLLALCQNOGKYEAVERY 354  
Qy 360 YORALAIYEGOLGPDNPVARTKNLASCYLKQGYAEATLYKEILTRAHVQEFSGVDD 419  
Db 355 YKRALEIYESKLGPDNPVARTKNLASCYLKQGYAEATLYKEILTRAHVQEFSGVDD 414  
Qy 420 DHKPIWMHAEEREMSKSRHHEGGTPYAEYGGWYKACKYSSPTVTTLNLGALYRROGK 479  
Db 415 ENKPIWQIAEERE---NKHKGEGATANEQAGWAKAAKVDSPVTVTTLKLNGLALYRROGK 471  
Qy 480 LEAETLEECALRSRQ-----GTDPISTQTKVAELLEGSDG---RTSQEG 522  
Db 472 YEAATLEEDVALRAKKQHEPLRSGAMGIDEMSQSMWASTIGGSRNSTTSTQ 527

RESULT 12  
S41865  
kinesin light chain (isoform 2) - Caenorhabditis elegans

C:Species: Caenorhabditis elegans  
C:Date: 06-Jan-1995 #sequence\_revision 06-Jan-1995 #text\_change 09-Sep-1997  
C:Accession: S41865  
R:Fan, J.; Amos, L.A.  
submitted to the EMBL Data Library, January 1994  
A:Description: Kinesin light chain isoforms in C. elegans.  
A:Reference number: S41864  
A:Accession: S41865  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-522 <FAN>  
A:Cross-references: EMBL:229645; NID:g452603; PID:g452604

Query Match 47.2%; Score 1491.5; DB 2; Length 522;  
Best Local Similarity 60.8%; Pred. No. 2.5e-79;  
Matches 302; Conservative 69; Mismatches 97; Indels 29; Gaps 7;

Qy 17 LSOEILGSTRVLSQGLEALRSEHQVLSQSLQTECLOQGGHEGLVHEKARQLRRSME 76  
Db 25 MSQDDVTGLRTVQOGLLEALREHSTISNTLETSVKGVKE--DEAPLPKQLSINDNLD 82  
Qy 77 NIELGLSEAQVMLALASHLSTVESEKQKRAQVRLRCQENQWLRLDELQAGTQORLQSEQA 136  
Db 83 KLVCGVDETSMLMVFLQTLQGMDAQHQYQAQRRRLCQENAWLRDELSSSTQIKLQOSEQM 142  
Qy 137 VAOLEEKKHLEFLGQLROYDEDEHTSEEKGG-----DATKDSLDDL-FPNEEED--- 186  
Db 143 VAOLEEENKHLKYMASIKQLD-DGTQSDTKTSVDVGPQVPTNETLQELGFGPEDEEDMNA 201  
Qy 187 -----PSNGLSRGOGATAAQGGYEIPARLTLNLVIOYAAQGYEYAVPLCKQALE 239  
Db 202 SQFNQPTPANQMA-----ASANVGEIPARLTLNLVIOYASQGYEYAVPLCKQALE 255  
Qy 240 DLERTSGRHPDVPATMLNIALVYRDQNKYKEAAHLNDALSTRESTLGPDPHPAVALTN 299  
Db 256 DLEKTSQGHDPDVPATMLNIALVYRDQNKYKEAAHLNDALSTRESTLGPDPHPAVALTN 315  
Qy 300 NLAVLYGRGKYKEAEPICQRALEIREKVLGTHNPDVAKQLNLLALCQNOGKYEAVERY 359  
Db 316 NLAVLYGRGKYKEAEPICQRALEIREKVLGTHNPDVAKQLNLLALCQNOGKYEAVERY 375  
Qy 360 YORALAIYEGOLGPDNPVARTKNLASCYLKQGYAEATLYKEILTRAHVQEFSGVDD 419  
Db 376 YKRALEIYESKLGPDNPVARTKNLASCYLKQGYAEATLYKEILTRAHVQEFSGVDD 435  
Qy 420 DHKPIWMHAEEREMSKSRHHEGGTPYAEYGGWYKACKYSSPTVTTLNLGALYRROGK 479  
Db 436 ENKPIWQIAEERE---NKHKGEGATANEQAGWAKAAKVDSPVTVTTLKLNGLALYRROGK 492  
Qy 480 LEAETLEECALRSRQ 496  
Db 493 YEAATLEEDVALRAKKQ 509

RESULT 13  
S33816  
kinesin light chain isoform 4 - sea urchin (Strongylocentrotus purpuratus)  
C:Species: Strongylocentrotus purpuratus (purple urchin)  
C:Date: 08-Dec-1993 #sequence\_revision 01-Dec-1995 #text\_change 09-Sep-1997  
C:Accession: S33816  
R:Wedaman, K.P.; Knight, A.E.; Kendrick-Jones, J.; Scholey, J.M.  
J. Mol. Biol. 231, 155-158, 1993  
A:Title: Sequences of sea urchin kinesin light chain isoforms.  
A:Reference number: S33813; MUID:93267648  
A:Accession: S33816  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-451 <WED>  
A:Cross-references: EMBL:L08258; NID:g161531; PID:g161532

Query Match 46.0%; Score 1453; DB 2; Length 451;



[illegible]

Qy	199	AAQ-----QGGEYIPARLTLNHLVLYAAQGRYEYAVPLCKQALEDLESTSGRHDPDA	253
		: : : : :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :	
Db	184	SSSFNPSINDYQLPTRLTQLNLVLYMEQGRFEVAIPCLCKQALEDVYKVGHNVLDA	243
		: : : : :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :	
Qy	254	TMLNIALVYRDONKYKEAAHLLNDALSTRESTLGDHPAVAATLNNLAVLYGKRKYKE	313
		: : : : :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :	
Db	244	TMLNVLIAVYRNGENFKDAIYLEKALSIRVQCCENHHSVAATLNNLAIAYGKRKYKE	303
		: : : : :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :	
Qy	314	AEPICQRALEIREKVLGTNHPDVAKOLNLLLCQNOGKYEAVERYORALAIYEGQLGP	373
		: : : : :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :	
Db	304	SEPUCKRALEIRKNLGNHPDVAKOLTWLGIVYQOLEKIETENYFKOALSIYNRAPPE	363
		: : : : :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :	
Qy	374	DNPVARTKNNLASCYLKGKQAEAEATLYKEILTRAHVQEGFSVDDDKHPIMWMAEERE	433
		: : : : :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :	
Db	364	NHQNVIKTKQLASVFLKGKQGEAEEMKNILSKVAI-----TGKNPTWRTAEDREE	416
		: : : : :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :	
Qy	434	MSKSRHREGTPTAAYEGGW-YRACKVSSPTVNTTLRNIGALYRROGKLEAAETLEECALR	492
		: : : : :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :	
Db	417	RQRN-----GIPKVDDESEFNVPNTVMDSNVNMTIKNLAAVYRKQKEEAAAGTLEE-ALG	470
		: : : : :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :	
Qy	493	SRQK---GTDPISQT	504
		: : : : :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :	
Db	471	AKQINGGADHTNST	485
		: : : : :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :	
		RESULT	15
		RS47998	
		kinesin light chain (isoform 2) - Caenorhabditis elegans	
		C/Species: Caenorhabditis elegans	
		C/Date: 07-May-1995 #sequence_revision 21-Jul-1995 #text_change 21-Jul-1995	
		C/Accession: S47998	
		R/Fan, J.; Amos, L.A.	
		J. Mol. Biol. 240, 507-512, 1994	
		A/Title: Kinesin light chain isoforms in Caenorhabditis elegans.	
		A/Reference number: S47997; MUID:94322393	
		A/Accession: S47998	
		A/Status: preliminary	
		A/Molecule type: mRNA	
		A/Residues: 1-114 <FAN>	
		A/Cross-references: EMBL:229645	

Search completed: August 15, 2000, 10:55:49  
Job time: 23076 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.  
OM protein - protein search, using sw model  
Run on: August 15, 2000, 10:52:23 ; Search time 42.09 Seconds  
(without alignments)  
455.744 Million cell updates/sec  
Title: US-09-036-614A-1  
Perfect score: 3161  
Sequence: 1 MSGVLGQRDEPAGRLSQE.....APLQVSRGLSASTMDLSSSS 619  
Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5  
Searched: 85661 seqs, 30989116 residues  
Total number of hits satisfying chosen parameters: 85661  
Minimum DB seq length: 0  
Maximum DB seq length: 1000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries  
Database : SwissProt\_38:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2000	63.3	569	1 KNL_C_HUMAN	Q07866 homo sapien
2	1985.5	62.8	556	1 KNL_C_RAT	P37285 rattus norv
3	1821	57.6	686	1 KNL_C_STRPU	Q05090 strongyloce
4	1759.5	55.7	571	1 KNL_C_LOLPE	P46825 loligo peal
5	1709	54.1	508	1 KNL_C_DROME	P46824 drosophila
6	1516	48.0	540	1 KNL_C_CAEEL	P46822 caenorhabdi
7	179.5	5.7	670	1 SR72_HUMAN	O76094 homo sapien
8	179	5.7	1966	1 MYSB_CAEEL	P02566 caenorhabdi
9	178.5	5.6	1321	1 IF3X_DICDI	O15818 dictyosteli
10	175.5	5.6	670	1 SR72_CANFA	P33731 canis famli
11	173.5	5.5	848	1 MYSB_DIRIM	P13392 dirofilaria
12	171.5	5.4	880	1 MYSB_BRUMA	Q01202 brugia mala
13	171.5	5.4	1857	1 MYSB_HUMAN	P35749 homo sapien
14	169.5	5.4	879	1 MYSB_ONCVO	Q02171 onchocerca
15	166	5.3	1972	1 MYSB_RABIT	P35748 oryctolagus
16	162.5	5.1	1300	1 DYNB_NEUPORA	Q01397 neurospora
17	162	5.1	1969	1 MYSB_CAEEL	P12844 caenorhabdi
18	160	5.1	1134	1 IF3X_HUMAN	O75153 homo sapien
19	160	5.1	1920	1 PONT_MOUSE	P48725 mus musculu
20	159.5	5.0	543	1 IF3S_HUMAN	P31948 homo sapien
21	158.5	5.0	3321	1 KEND_HUMAN	O05613 homo sapien
22	157.5	5.0	866	1 MYSB_SCHMA	P06198 schistosoma
23	155.5	4.9	1036	1 OGT1_RAT	P36558 rattus norv
24	155.5	4.9	1505	1 CDP_HUMAN	P39880 homo sapien
25	154	4.9	619	1 OM70_NEUCR	P23231 neurospora
26	153.5	4.9	882	1 MYSB_CAEEL	P10567 caenorhabdi
27	153	4.8	1938	1 MYSB_CAEEL	P02567 caenorhabdi
28	152.5	4.8	1325	1 G60_MOUSE	P55937 mus musculu
29	152	4.8	820	1 FES_FELCA	P14238 felis silve
30	151.5	4.8	920	1 OGT1_HUMAN	O15294 homo sapien
31	151.5	4.8	1960	1 MYSB_HUMAN	P35579 homo sapien
32	151	4.8	576	1 PEX5_PICPA	P33292 pichia past
33	150.5	4.8	1597	1 CTRO_MOUSE	P49025 mus musculu

RESULT 1

KNLC\_HUMAN

ID KNLC\_HUMAN

AC Q07866;

DT 01-OCT-1994 (Rel. 30, Created)

DT 01-OCT-1994 (Rel. 30, Last sequence update)

DT 01-OCT-1996 (Rel. 34, Last annotation update)

DE KINESIN LIGHT CHAIN (KLC).

GN KLC1 OR KLC OR KNS2.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE; 94099888.

RA Cabeza-Arvelaiz Y., Shih L.-C.N., Hardman N., Asselbergs F.,

RA Bilbe G., Schmitz A., White B., Siciliano M.J., Lachman L.B.;

RT "Cloning and genetic characterization of the human kinesin

RT light-chain (KLC) gene".

RL DNA Cell Biol. 12:881-892(1993).

CC -1- FUNCTION: KINESIN IS A MICROTUBULE-ASSOCIATED FORCE-PRODUCING

CC PROTEIN THAT MAY PLAY A ROLE IN ORGANELLE TRANSPORT. THE LIGHT

CC CHAIN MAY FUNCTION IN COUPLING OF CARGO TO THE HEAVY CHAIN OR IN

CC THE MODULATION OF ITS ATPASE ACTIVITY.

CC -1- SUBUNIT: OLIGOMERIC COMPLEX COMPOSED OF TWO HEAVY CHAINS AND

CC TWO LIGHT CHAINS.

CC -1- ALTERNATIVE PRODUCTS: AS FOR THE RAT KINESIN LIGHT CHAIN, THE

CC HUMAN PRODUCT MAY WELL UNDERGO ALTERNATIVE SPLICING.

CC -1- TISSUE SPECIFICITY: FOUND IN A VARIETY OF TISSUES. MOSTLY ABUNDANT

CC IN BRAIN AND SPINE.

CC -1- DOMAIN: THE LIGHT CHAIN IS COMPOSED OF THREE STRUCTURAL DOMAINS: A

CC LARGE GLOBULAR N-TERMINAL DOMAIN WHICH MAY BE INVOLVED IN BINDING

CC TO KINESIN HEAVY CHAINS, A CENTRAL ALPHA-HELICAL COILED-COIL

CC DOMAIN THAT MEDIATES THE LIGHT CHAIN DIMERIZATION; AND A SMALL

CC GLOBULAR C-TERMINAL WHICH MAY PLAY A ROLE IN REGULATING

CC MECHANOCHEMICAL ACTIVITY OR ATTACHMENT OF KINESIN TO MEMBRANE-

CC BOUND ORGANELLES.

CC -1- SIMILARITY: BELONGS TO THE KINESIN LIGHT CHAIN FAMILY.

CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/) or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

CC EMBL; L04733; AAA16576.1; -

CC MIM; 600025; -

CC PFAM; PF00409; kinesin2; 5.

CC PRINTS; PR00381; KINESINLIGHT.

CC PROSITE; PS01160; KINESINLIGHT; 4.

CC Motor protein; Microtubules; Coiled coil; Repeat;

CC Alternative splicing.

CC DOMAIN 45 150 COILED COIL.

CC DOMAIN 234 401 4 X 42 AA APPROXIMATE TANDEM REPEATS.

34	149.5	4.7	955	1	KINL_LEICH	P46865 leishmania
35	149.5	4.7	1805	1	NEST_RAT	P21263 rattus norv
36	149	4.7	1947	1	MYS_CAEEL	P12845 caenorhabdi
37	148.5	4.7	1938	1	MYS_AEQIR	P24733 aequipecten
38	148	4.7	1509	1	MYSN_ACACA	P05659 acanthamoeb
39	147	4.7	3210	1	CENF_HUMAN	P49454 homo sapien
40	146.5	4.6	866	1	MYSB_SCHJA	Q05870 schistosoma
41	146.5	4.6	2411	1	MYSB_DROME	P05661 drosophila
42	146	4.6	1978	1	MYSQ_CHICK	P10587 gallus gall
43	145.5	4.6	1140	1	MYSB_HUMAN	P11055 homo sapien
44	144.5	4.6	1155	1	IF3X_SCHPO	O59742 schizosacch
45	144	4.6	1928	1	MYSI_YEAST	P08964 saccharomyc

ALIGNMENTS

```

FT REPEAT      234      275      1.
FT REPEAT      276      317      2.
FT REPEAT      318      359      3.
FT REPEAT      360      401      4.
SQ SEQUENCE    569 AA; 64786 MW; 5D9376C3EEDD00FE CRC64;

Query Match      63.3%; Score 2000; DB 1; Length 569;
Best Local Similarity 69.8%; Pred. No. 3.5e-105;
Matches 395; Conservative 77; Mismatches 76; Indels 18; Gaps 8;

QY 1 MSGVLVQORDEPAGHRLSOFELIGSTRVYSGLEALRSEHQAVLSQTSICLQGGHE 60
DB 1 MSTNVIYIKEDKL--EKTDQDEITSKTKQVIOGLEALKNHNSLSLLETLCKLR-KDDE 57

QY 61 EGVHEKAROLRSMENIELGLSEAOVMALASHLSTVESEKOKLRAQVRRRCQENQWL 120
DB 58 SNLVEEKNMIRKSLMLELGLSEAOVMALSNHNAVESEKOKLRAQVRRRCQENQWL 117

QY 121 DELAGTORLQORSEQAVALKEEKKHLEFLGQLRQYDEGHTSEKEGDATKDSLDLDP 180
DB 118 DELANTQOKLKSEQSAVALKEEKKHLEFMNQLKDYDDDISPSEDKDTSTREPLDLP 177

QY 181 NEPEDEPSNGLSR--GOGATAOOGGVEIPARLTLNHLNVIQYAAOGRYEAVPLCKQAL 238
DB 178 N-DEDDPGQIQOQHSHSAAAQOOGGVEIPARLTLNHLNVIQYASOGRYEAVPLCKQAL 236

QY 239 EDLERTSGRHPDVATMLNIALVYRDQNKYKEAAHLLNDALSIRESTLGPDPHVAATL 298
DB 237 EDLEKTSGHDPDVATMLNIALVYRDQNKYKDAANLLNDALAIKTLGKHPHVAATL 296

QY 299 NNLAVALYKGRGKYKEAPLQORALEIREKVLGTHNHPDVAKQNNLALLCONQKGYEAV 358
DB 297 NNLAVALYKGRGKYKEAPLQORALEIREKVLGTHNHPDVAKQNNLALLCONQKGYEAV 356

QY 359 YQORALAIYEGQLGPDNPVARTKNNLASCYLKQGYAEATLYKELTRAHVQEGSYD 418
DB 357 YQORALAIYEGQLGPDNPVARTKNNLASCYLKQGYAEATLYKELTRAHVQEGSYD 416

QY 419 DDHKPIWHAEREEMSKSRHEGGTPYAEYGGWKYACKYSSPTVNTTLNGLALYRROG 478
DB 417 DENKPIWHAEREEMSKSRHEGGTPYAEYGGWKYACKYSSPTVNTTLNGLALYRROG 474

QY 479 KLEAAETLECALRSRQGTDPISQTKVABLLGESDG---RRTSQGPGDSYKFE----G 531
DB 475 KFEAAETLEAAARSRQGLDNVHKQYAEVLNDPENMEKRERRSLNVDVWVYKESGPDG 534

QY 532 GEDASVAVESGDSGTLQORSGISGK 557
DB 535 GEEVSMVSEWNGVSG---RASFCGR 557

RESULT 2
KNLC_RAT
ID KNLC_RAT STANDARD; PRT; 556 AA.
AC P37285;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE KINESIN LIGHT CHAIN (KLC).
GN KLC1 OR KLC.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN;
RC MEDLINE; 92052221.
RA Cyr J.L., Pfister K.K., Bloom G.S., Slaughter C.A., Brady S.T.;
RT "Molecular genetics of kinesin light chains: generation of isoforms
RT by alternative splicing."
RL Proc. Natl. Acad. Sci. U.S.A. 88:10114-10118(1991).
CC -1- FUNCTION: KINESIN IS A MICROTUBULE-ASSOCIATED FORCE-PRODUCING

```

```

CC CC PROTEIN THAT MAY PLAY A ROLE IN ORGANELLE TRANSPORT. THE LIGHT
CC CC CHAIN MAY FUNCTION IN COUPLING OF CARGO TO THE HEAVY CHAIN OR IN
CC CC THE MODULATION OF ITS ATPASE ACTIVITY.
CC CC -1- SUBUNIT: OLIGOMERIC COMPLEX COMPOSED OF TWO HEAVY CHAINS AND
CC CC TWO LIGHT CHAINS.
CC CC -1- ALTERNATIVE PRODUCTS: AT LEAST THREE FORMS (A,B AND C) ARE
CC CC PRODUCED BY ALTERNATIVE SPLICING. THEY ONLY DIFFER IN THE C-
CC CC TERMINUS. THE SEQUENCE OF FORM C IS SHOWN HERE.
CC CC -1- DOMAIN: THE LIGHT CHAIN IS COMPOSED OF THREE STRUCTURAL DOMAINS: A
CC CC LARGE GLOBULAR N-TERMINAL DOMAIN WHICH MAY BE INVOLVED IN BINDING
CC CC TO KINESIN HEAVY CHAINS, A CENTRAL ALPHA-HELICAL COILED-COIL
CC CC DOMAIN THAT MEDIATES THE LIGHT CHAIN DIMERIZATION; AND A SMALL
CC CC GLOBULAR C-TERMINAL WHICH MAY PLAY A ROLE IN REGULATING
CC CC MECHANOCHEMICAL ACTIVITY OR ATTACHMENT OF KINESIN TO MEMBRANE-
CC CC BOUND ORGANELLES.
CC CC -1- SIMILARITY: BELONGS TO THE KINESIN LIGHT CHAIN FAMILY.
CC CC
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC
CC CC EMBL; M75148; -; NOT_ANNOTATED_CDS.
CC CC EMBL; M75147; -; NOT_ANNOTATED_CDS.
CC CC EMBL; M75146; -; NOT_ANNOTATED_CDS.
CC CC PFAM; PF00409; Kinesin2; 5.
CC CC PRINTS; PR00381; KINESINLIGHT.
CC CC PROSITE; PS01160; KINESIN_LIGHT; 4.
CC CC Motor protein; Microtubules; Coiled coil; Repeat;
CC CC Alternative splicing.
CC CC DOMAIN 45 150 COILED COIL.
CC CC FT DOMAIN 234 401 4 X 42 AA APPROXIMATE TANDEM REPEATS.
CC CC FT REPEAT 234 275 1.
CC CC FT REPEAT 276 317 2.
CC CC FT REPEAT 318 359 3.
CC CC FT REPEAT 360 401 4.
CC CC FT VARSPLIC 538 546 MISSING (IN ISOFORM B).
CC CC FT VARSPLIC 538 556 VSMVSEWNGMKMLGLVK -> A (IN ISOFORM A).
CC CC SQ SEQUENCE 556 AA; 63247 MW; 1A843B34BCDDB53E CRC64;

Query Match      62.8%; Score 1985.5; DB 1; Length 556;
Best Local Similarity 70.2%; Pred. No. 2.2e-104;
Matches 388; Conservative 76; Mismatches 70; Indels 19; Gaps 7;

QY 16 RLSQEEILGSTRVYSGLEALRSEHQAVLSQTSICLQGGHEGLVHEKARQLRSM 75
DB 14 KLTQDEITSKTKQVIOGLEALKNHNSLSLLETLCKLR-KDDESNLVEKSMIRKSL 72

QY 76 ENIELGLSEAOVMALASHLSTVESEKOKLRAQVRRRCQENQWLRLDELQTLQORLSEQ 135
DB 73 EMLELGLSEAOVMALSNHNAVESEKOKLRAQVRRRCQENQWLRLDELQTLQORLSEQ 132

QY 136 AVAQLEEEKKHLEFLGQLRQYDEGHTSEKEGDATKDSLDLDPNEEEDPSNGLSR-- 193
DB 133 SVAQLEEEKKHLEFMNQLKDYDDDISPSEDKDSKSEPLDLPN-DEDDPGQIQOQH 191

QY 194 QGATAOOGGVEIPARLTLNHLNVIQYAAOGRYEAVPLCKQALEDLRTSGRHPDVA 253
DB 192 SSSAAAAOOGGVEIPARLTLNHLNVIQYASOGRYEAVPLCKQALEDLRTSGRHPDVA 251

QY 254 TMLNIALVYRDQNKYKEAAHLLNDALSIRESTLGPDPHVAATLNNLALYKGRGYKE 313
DB 252 TMLNIALVYRDQNKYKDAANLLNDALAIKTLGDRHPHVAATLNNLALYKGRGYKE 311

QY 314 AEPLQORALEIREKVLGTHNHPDVAKQNNLALLCONQKGYEAVERYQORALAIYEGQLP 373
DB 312 AEPLCKRALEIREKVLGTHNHPDVAKQNNLALLCONQKGYEVEEYQORALEIYQTLGP 371

QY 374 DNPVARTKNNLASCYLKQGYAEATLYKELTRAHVQEGSYDDDKHPIWHAERE 433

```

Db 372 DPNVAKTANNLASCYLKQKFKQKQETLYKEILTRAHERFSGVDENKPLWHAERE 431  
QY 434 MSKSRHEGGTPYAEYGGYKACKYSSPTVNTTLRNALGYRRQKLEAAETLECALRS 493  
Db 432 CRGKQ--KDGSGFEGYGGYKACKYSDPTVTTTLNKGALYRRQKFEAAETLEAALRS 489  
QY 494 RQGGTDPISTQKVAELLGSDG---RRTSQEGPGDSVKFE---GGEDASVAVESWGDGS 546  
Db 490 RKGQDNVHKQVAEYLVNDPENVERKRSRLNVDVRYESGPDGGEVMSVEMNG--- 546  
QY 547 GILQSGSLGKIR 559  
Db 547 ---MRKMKLGIVK 556

RESULT 3  
KNLC\_STRPU STANDARD; PRT; 686 AA.  
AC Q05090; Q05089; Q05088; Q04801;  
DT 01-JUN-1994 (Rel. 29, Created)  
DT 01-JUN-1994 (Rel. 29, Last sequence update)  
DT 15-FEB-2000 (Rel. 39, Last annotation update)  
DE KINESIN LIGHT CHAIN (KLC).  
OS Strongylocentrotus purpuratus (Purple sea urchin).  
OC Eukaryota; Metazoa; Echinodermata; Echinozoa; Echinoidea;  
OC Euechinoidea; Echinacea; Echinoida; Strongylocentrotidae;  
OC Strongylocentrotus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=EGG;  
RX MEDLINE; 93267648.  
RA Wedman K.P.; Knight A.E.; Kendrick-Jones J.; Scholey J.M.;  
RT "Sequences of sea urchin kinesin light chain isoforms.";  
RL J. Mol. Biol. 231:155-158(1993).  
CC -!- FUNCTION: KINESIN IS A MICROTUBULE-ASSOCIATED FORCE-PRODUCING  
CC PROTEIN THAT MAY PLAY A ROLE IN ORGANELLE TRANSPORT. THE LIGHT  
CC CHAIN MAY FUNCTION IN COUPLING OF CARGO TO THE HEAVY CHAIN OR IN  
CC THE MODULATION OF ITS ATPASE ACTIVITY.  
CC -!- SUBUNIT: OLIGOMERIC COMPLEX COMPOSED OF TWO HEAVY CHAINS AND  
CC TWO LIGHT CHAINS.  
CC -!- ALTERNATIVE PRODUCTS: AT LEAST FOUR FORMS (KLC-1 TO KLC-4) ARE  
CC PRODUCED BY ALTERNATIVE SPLICING. THE SEQUENCE OF FORM 3 (KLC-3)  
CC IS SHOWN HERE.  
CC -!- DOMAIN: THE LIGHT CHAIN IS COMPOSED OF THREE STRUCTURAL DOMAINS: A  
CC LARGE GLOBULAR N-TERMINAL DOMAIN WHICH MAY BE INVOLVED IN BINDING  
CC TO KINESIN HEAVY CHAINS, A CENTRAL ALPHA-HELICAL COILED-COIL  
CC DOMAIN THAT MEDIATES THE LIGHT CHAIN DIMERIZATION; AND A SMALL  
CC GLOBULAR C-TERMINAL WHICH MAY PLAY A ROLE IN REGULATING  
CC MECHANOCHEMICAL ACTIVITY OR ATTACHMENT OF KINESIN TO MEMBRANE-  
CC BOUND ORGANELLES.  
CC -!- PTM: PHOSPHORYLATION MAY MODULATE THE PROCESS OF MECHANOCHEMICAL  
CC COUPLING.  
CC -!- SIMILARITY: BELONGS TO THE KINESIN LIGHT CHAIN FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; L10235; AAA03059.1; -  
CC EMBL; L10234; AAA03058.1; -  
CC EMBL; L10233; AAA03057.1; -  
CC EMBL; L08258; AAA03060.1; -  
CC PIR; S33814; S33814.  
CC PIR; S33813; S33813.  
CC PIR; S33816; S33816.  
CC PFAM; PF00409; kinesin2; 5.  
CC PRINTS; PR00381; KINESINLIGHT.  
CC PROSITE; PS01160; KINESIN\_LIGHT; 4.

KW Motor protein; Microtubules; Coiled coil; Repeat;  
KW Alternative splicing; Phosphorylation.  
FT DOMAIN 20 160 COILED COIL.  
FT REPEAT 240 407 4 X 42 AA APPROXIMATE TANDEM REPEATS.  
FT REPEAT 280 323 1.  
FT REPEAT 324 365 2.  
FT REPEAT 366 407 3.  
FT VARSPLIC 564 600 MISSING (IN ISOFORM KLC-2).  
FT VARSPLIC 564 600 MISSING (IN ISOFORM KLC-1).  
FT VARSPLIC 441 451 GFKDNAPYGD -> VKRKPKPAKS (IN ISOFORM  
FT VARSPLIC 452 451 MISSING (IN ISOFORM KLC-4).  
FT VARSPLIC 452 451 MISSING (IN ISOFORM KLC-4).  
SQ SEQUENCE 686 AA; 76517 MW; 603D71186CC0364B CRC64;

Query Match 57.6%; Score 1821; DB 1; Length 686;  
Best Local Similarity 57.2%; Pred. No. 4.6e-95;  
Matches 395; Conservative 82; Mismatches 129; Indels 84; Gaps 16;  
QY 1 MSGVLGQRDEPAGHR--LSQEEILGSTRVLSQGLALRSEHQAVLQSLQSTIECLQQG- 57  
Db 1 MSGKSLSTPNNSGGQGNLSQEOIITGTREVIGLQLEKNEHDILNSLYQSLKMLKKT 60  
QY 58 GHEGLVHEKARQLRRSMENIELGLSEAQVMLALASHLSTVESEKOKLRAQVRLQCEQ 117  
Db 61 PGDSNLVEEKTDIIEKSLSELEGLGEAKVMAALGHNLNVEAEKQKLAQVRRVQENT 120  
QY 118 WLRLDELAGTQORLQSEQAQLEEEKHLEFLGQLROYDEGDGHTSEEGEGATKDSLD 177  
Db 121 WLRLDLAATQOKLOTSEQNLADLEVYKHKLEYMNSIKYDEDTDPDEEA---SSDPLD 177  
QY 178 LFPNEEEDPSNGLS-----RGOGATAAQOGGYEIPARLTLNLNLTQVAAQRYEV 229  
Db 178 GFP-----EDDDGGQADESYPPQTGSGVSAAGGYEIPARLTLNLNLTQVAAQRYEV 233  
QY 230 AVPLCKQALDELRTSGRHPDVATMLNLTALVYRDQNKYKAAHLNLDALSIRETLGP 289  
Db 234 AVPLCKQALDELRTSGRHPDVATMLNLTALVYRDQNKYKAAHLNLDALSIRETLGP 293  
QY 290 DHPAVALTNLNLAVLYGKRGKYKEAPLCQRALEIREKVLGTNHPDVAKQLNLLCON 349  
Db 294 DHPAVALTNLNLAVLYGKRGKYKEAPLCQRALEIREKVLGTNHPDVAKQLNLLCON 353  
QY 350 QGYEAVERYQALAIYEGQLGPDNPVARTKNNLASCYLKQGYAEATLYKEITLTRA 409  
Db 354 QGYEAVERYQALAIYEGQLGPDNPVARTKNNLASCYLKQGYAEATLYKEITLTRA 413  
QY 410 HVQEFQ-SVDD-DHKPIWMHAEREEMSKSRHEGGTTPYAEYGGYKACKV-----SSPT 462  
Db 414 HERFGLSADDKDKNPIMQAEREEREKGFKN---APYDYGWGHKAQVDSRSRSP 470  
QY 463 VNTTLNLGALYRRQKLEAAETLECALRSRQGTDPISQTKVAELLG----- 511  
Db 471 VNTTLNLGALYRRQKLEAAETLECALRSRQGTDPISQTKVAELLG----- 530  
QY 512 -----ESDGRTS---QEGPDSDVKFE---GGEDASVAVESWGDGS 543  
Db 531 AMAKERHRRSSCTPHRGSTESVYKTDGSEVSGVAVAKKAKKAKKAKKAKKAKKAKKAKK 590  
QY 544 -----DGSGLTQSGSLGKIRDLVLRSSSELVLRKLOGTETPRPSSNNKRAASLN 593  
Db 591 PRSPPHVLVNGDKLRRSGSLKLRASVRSSTKLNLKLGRE-SDDDDGGMKRASSMSV 649  
QY 594 L-----NQPSAAPLOYS-RGLSASTMDLSS 617  
Db 650 LPSRGNDSTPAPIQLSQRGRVSGHDNLSS 679  
RESULT 4  
KNLC\_LOLPE STANDARD; PRT; 571 AA.  
ID KNLC\_LOLPE  
AC P46825;



TO KINESIN HEAVY CHAINS, A CENTRAL ALPHA-HELICAL COILED-COIL DOMAIN THAT MEDIATES THE LIGHT CHAIN DIMERIZATION; AND A SMALL GLOBULAR C-TERMINAL WHICH MAY PLAY A ROLE IN REGULATING MECHANOCHEMICAL ACTIVITY OR ATTACHMENT OF KINESIN TO MEMBRANE-BOUND ORGANELLES.

-!- SIMILARITY: BELONGS TO THE KINESIN LIGHT CHAIN FAMILY.

THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL: L11013; AAA02481.1; -;  
EMBL: L11328; AAA28669.1; -;  
FLYBASE; FBgn010235; Klc.  
PFAM; PF00409; kinesin2; 4.  
PRINTS; PR00381; KINESINLIGHT.  
DR PROSITE; PS01160; KINESINLIGHT; 4.  
KW MOTOR protein; Microtubules; Coiled coil; Repeat.  
FT DOMAIN 34 129  
FT DOMAIN 211 378  
FT REPEAT 211 252 1.  
FT REPEAT 253 294 2.  
FT REPEAT 295 336 3.  
FT REPEAT 337 378 4.  
SQ SEQUENCE 508 AA; 58044 MW; B7893FF47EB638A9 CRC64;

Query Match 54.13; Score 1709; DB 1; Length 508;  
Best Local Similarity 67.28; Pred. No. 5.8e-99;  
Matches 338; Conservative 66; Mismatches 85; Indels 14; Gaps 5;  
Qy 16 RLSEEEILGSTRVUSQGLEALRSEHQAVLSQITIECLOQGGHEEGLVHEKARQLRSM 75  
Db 3 QMSQDEIITNTKTVLQGLEALRSEHVSVIMNGIAE---VQKD-----NEKSDMLRKN 51

Qy 76 ENIELGSEAOVMALASHLSTVESEKURAVRRICQENQWRDLACTQRLQSRSEQ 135  
Db 52 ENIELGSEAOVMALASHLSTVESEKURAVRRICQENQWRDLACTQRLQSRSEQ 111  
Qy 136 AVAQLEEEKHLEFLGLQRYDEGHTSEEKGDATKDSLDLFPNEEEDPSNLSRGQ 195  
Db 112 LVAQLEEEKHLEFMAWVKYDENOEDDADCKSRT-DPVVEFLPPDENEDRNHNSPTTP 170  
Qy 196 GATAAQGGVEIPARLRLHNLVIOYAAQGRYEVAVPLCKQALDLERTSGRHPDVATM 255  
Db 171 SQFANQTSYGEIPARLRLHNLVIOYAAQGRYEVAVPLCKQALDLERTSGRHPDVATM 230  
Qy 256 LNLIALYRDOVKYKEAHLNDALSIRESTLGPDPHFAVAATLNNLAALYGRGKYKEAE 315  
Db 231 LNLIALYRDOVKYKEAHLNDALSIRGTLGHNFAVAATLNNLAALYGRGKYKDAE 290  
Qy 316 PLCQRALEIREKVLGTHNPDVAKQNNLALICQHQGRYEAVERVYQALAIYEGQLGPDN 375  
Db 291 PLCRALEIREKVLGTHNPDVAKQNNLALICQHQGRYEAVERVYQALAIYEGQLGPDN 350  
Qy 376 PNVAARTNNLASCYLKQKGYAEATLYKEITLTRAHQVQFGSDDDHKPIWMAHEEREEMS 435  
Db 351 PNVAARTNNLASCYLKQKGYAEATLYKEITLTRAHQVQFGSDDDHKPIWMAHEEREEMS 409  
Qy 436 KSRHHEGGTPVAEYGGVYKACKYSSPVNTLNLGALYRQGLKLEAETLECALRSRR 495  
Db 410 KFDNRE-NTPFYGGVGHKAAKVDSPVTTTLNLGALYRQGMFEAETLEDCAMRSKK 468  
Qy 496 QGTDPIQSQTVAELLGSDGRRT 518  
Db 469 EAYDLAKQTKLSQLLTSNEKRRS 491

RESULT 6

KNLC\_CABEL STANDARD; PRT; 540 AA.  
ID KNLC\_CABEL  
AC P46822; Q18088;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE KINESIN LIGHT CHAIN (KLC).  
GN KLC-2 OR C18C4.10.  
OS Caenorhabditis elegans.  
OC Rhabdityda; Metazoa; Nematoda; Chromadorea; Rhabdityda; Rhabdityda;  
OC Rhabdityda; Peloderinae; Caenorhabditis.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 94322393.  
RA Fan J., Amos L.A.;  
RT "Kinesin light chain isoforms in Caenorhabditis elegans."  
RL J. Mol. Biol. 240:507-512(1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RA Gattung S.;  
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: KINESIN IS A MICROTUBULE-ASSOCIATED FORCE-PRODUCING  
CC PROTEIN THAT MAY PLAY A ROLE IN ORGANELLE TRANSPORT. THE LIGHT  
CC CHAIN MAY FUNCTION IN COUPLING OF CARGO TO THE HEAVY CHAIN OR IN  
CC THE MODULATION OF ITS ATPASE ACTIVITY.  
CC -!- SUBUNIT: OLIGOMERIC COMPLEX COMPOSED OF TWO HEAVY CHAINS AND  
CC TWO LIGHT CHAINS.  
CC -!- ALTERNATIVE PRODUCTS: AT LEAST TWO FORMS ARE PRODUCED BY  
CC ALTERNATIVE SPLICING.  
CC -!- DOMAIN: THE LIGHT CHAIN IS COMPOSED OF THREE STRUCTURAL DOMAINS: A  
CC LARGE GLOBULAR N-TERMINAL DOMAIN WHICH MAY BE INVOLVED IN BINDING  
CC TO KINESIN HEAVY CHAINS, A CENTRAL ALPHA-HELICAL COILED-COIL  
CC DOMAIN THAT MEDIATES THE LIGHT CHAIN DIMERIZATION; AND A SMALL  
CC GLOBULAR C-TERMINAL WHICH MAY PLAY A ROLE IN REGULATING  
CC MECHANOCHEMICAL ACTIVITY OR ATTACHMENT OF KINESIN TO MEMBRANE-  
CC BOUND ORGANELLES.  
CC -!- SIMILARITY: BELONGS TO THE KINESIN LIGHT CHAIN FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; 229644; CA82752.1; -;  
EMBL; 229645; CA82753.1; -;  
EMBL; U55369; AAC25830.1; ALT\_INIT.  
DR WORKPEP; C18C4.10; CE08655.  
DR PFAM; PF00409; kinesin2; 5.  
DR PRINTS; PR00381; KINESINLIGHT.  
DR PROSITE; PS01160; KINESINLIGHT; 4.  
KW Motor protein; Microtubules; Coiled coil; Repeat;  
KW Alternative splicing.  
FT DOMAIN 34 138  
FT DOMAIN 231 398  
FT REPEAT 231 272 1.  
FT REPEAT 273 314 2.  
FT REPEAT 315 356 3.  
FT REPEAT 357 398 4.  
FT VARSPLIC 1 1  
FT VARSPLIC 489 540  
FT VARSPLIC 375 375  
FT CONFLICT 141 141  
FT CONFLICT L -> N (IN CA82753).  
FT CONFLICT L -> F (IN REF. 2).  
SQ SEQUENCE 540 AA; 60190 MW; 642C1D046557C34D CRC64;

Query Match 48.08; Score 1516; DB 1; Length 540;





P02566;  
 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 15-JUL-1998 (Rel. 36, Last annotation update)  
 DE MYOSIN HEAVY CHAIN B (MHC B).  
 GN UNC-54 OR MYO-4.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Peleoderinae; Caenorhabditis.  
 [1]  
 RN SEQUENCE FROM N.A.  
 RP MEDLINE; 83273600.  
 RA Karn J., Brenner S., Barnett L.;  
 RT "Protein structural domains in the Caenorhabditis elegans unc-54  
 RT myosin heavy chain gene are not separated by introns.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 80:4253-4257(1983).  
 [2]  
 RN SEQUENCE OF 850-1966 FROM N.A.  
 RP MEDLINE; 82272395.  
 RA McLachlan A.D., Karn J.;  
 RT "Periodic charge distributions in the myosin rod amino acid sequence  
 RT match cross-bridge spacings in muscle.";  
 RL Nature 299:226-231(1982).  
 [3]  
 RN SEQUENCE OF 1876-1966 FROM N.A.  
 RP MEDLINE; 83232892.  
 RA Wills N., Gesteland R.F., Karn J., Barnett L., Bolten S.,  
 RA Waterston R.H.;  
 RT "The genes sup-7 X and sup-5 III of C. elegans suppress amber  
 RT nonsense mutations via altered transfer RNA.";  
 RL Cell 33:575-583(1983).  
 CC -1- FUNCTION: MUSCLE CONTRACTION.  
 CC -1- SUBUNIT: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2  
 CC HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC)  
 CC AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).  
 CC -1- SUBCELLULAR LOCATION: THICK FILAMENTS OF THE MYOFIBRILS.  
 CC -1- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING  
 CC CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,  
 CC CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.  
 CC -1- PTM: TWO CYSTEINE RESIDUES IN THE S1 DOMAIN ARE SELECTIVELY  
 CC ALKYLATED AND ARE REQUIRED FOR MYOSIN ATPASE ACTIVITY.  
 CC -1- MISCELLANEOUS: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT  
 CC MEROMYOSIN (LMW) AND 1 HEAVY MEROMYOSIN (HMW). IT CAN LATER BE  
 CC SPLIT FURTHER INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED  
 CC SUBFRAGMENT (S2).  
 CC -1- MISCELLANEOUS: THERE ARE FOUR DIFFERENT MYOSIN HEAVY CHAINS IN C.  
 CC ELEGANS.  
 CC -1- MISCELLANEOUS: MHC A AND MHC B ARE FOUND EXCLUSIVELY IN THE BODY  
 CC WALL MUSCLE. THEY CO-ASSEMBLE INTO BODY WALL THICK FILAMENT.  
 CC -1- SIMILARITY: THE GLOBULAR HEAD AA SEQUENCE SHOWS A HIGH DEGREE OF  
 CC SIMILARITY WITH THE GLOBULAR HEAD SEQUENCES OF MUSCLE & NONMUSCLE  
 CC HEAVY CHAINS. BY CONTRAST THE ROD SEQUENCE IS LESS CONSERVED, BUT  
 CC THE PERIODICITIES OF HYDROPHOBIC & CHARGED RESIDUES, WHICH DICTATE  
 CC THE ALPHA-HELICAL COILED-COIL STRUCTURE, ARE CONSERVED.  
 CC -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; J01050; AAA28124.1; -;  
 DR EMBL; V01494; CAA24738.1; -;  
 DR PIR; A02992; MKW.  
 DR HSP; P08799; INMD.  
 DR PFM; P01576; Myosin\_tail; 1.  
 DR PFM; P00063; myosin\_head; 1.  
 DR PRINTS; PR00193; MYOSINHEAVY.  
 KW Myosin; Muscle protein; Coiled coil;  
 KW ATP-binding; Methylation; Alkylation; Heptad repeat pattern;  
 KW Multigene family.

FT	DOMAIN	1	850	GLOBULAR HEAD (S1).
FT	DOMAIN	851	1966	RODLIKE TAIL (S2 AND LMM DOMAINS).
FT	DOMAIN	851	1164	ALPHA-HELICAL TAILPIECE (S2).
FT	DOMAIN	1165	1966	LIGHT MEROMYOSIN (LMM).
FT	DOMAIN	851	1966	COILED COIL (POTENTIAL).
FT	NP_BIND	177	184	ATP (BY SIMILARITY).
FT	DOMAIN	665	687	ACTIN-BINDING.
FT	DOMAIN	769	783	ACTIN-BINDING.
FT	MOD_RES	128	128	METHYLATION (TRI-) (POTENTIAL).
FT	MOD_RES	705	705	ALKYLATION (SH-1).
FT	MOD_RES	715	715	ALKYLATION (SH-2).
FT	DOMAIN	1165	1176	HINGE.
FT	CONFLICT	1337	1337	E -> R (IN REF. 2).
FT	CONFLICT	1880	1880	I -> L (IN REF. 2).
SQ	SEQUENCE	1966	AA; 225125 MW; B66F0B2F27B67F CRC64;	

Query Match 5.7%; Score 179; DB 1; Length 1966;  
 Best Local Similarity 22.1%; Pred. No. 0.013;  
 Matches 158; Conservative 120; Mismatches 254; Indels 184; Gaps 36;

QY	30	SQGLEALRSEHOAVLSQSLQTEICLQGGHEEG--LVHEKARQLRRSMENIELGLSEAQV	87
DB	858	AELEKINDKVALEDLSLAKE-EKURKELEESSAKLVEEKT-SLFTNLESTKTOLSDAE	915
QY	88	MLA-----LASH-----LSTVESEKOKRAQV-----RR	111
DB	916	RLAKLEAQKQDASKQLSELNDQADNEDRTADVORAKKIEAEVEALKQIODLEMSLRK	975
QY	112	LCQENWLDELACTQORLQSENAVAOLEBEKKHLEFLGQLQYDDGDGTSEKEDAT	171
DB	976	ASEKSKDQHQIRSLQDEMQQOQDEAIAKLNEKKHQBIEIN--RKLMDLOSEEDKQNEQN	1033
QY	172	K-----DSLDLDFNEEE-----	192
DB	1034	KYKAKLEQLDLDLSELEKRRARADLDKQKRVGELKIAQENIDESGRQRHLENNLK	1093
QY	193	RGQGATAAOGGYEYAPARLTNLNVIQYAAQ---GRYEVAVPLCKQALDEL	243
DB	1094	KKESEL-----HSVSRLEDEQALVSKLQKIQDQGRIS-----ELEELENEROSRS	1142
QY	244	TSRGHPDVATMLNIALVYRQ-----NKYKEA---AHL---LNDALSIRESTLG	288
DB	1143	KADRAKSDIQLRELEELGELKDEGGATAAQVEVNNKRAELAKLRDLLEANNHNNQLG	1202
QY	289	----PHPVAATLNNLAVLYOKRGK-----YKEAEPLCQRALEIREKVLG-TNHPD	335
DB	1203	GLRKKHTDAVAELTDQLDQLNKAKAKVERKRAQAVRDAEDL---AAQLDQSTGKLNNEK	1259
QY	336	VAKQNLNALLCQNGKYEAVRYQORALAIYEGOLGPDNPVARTKNNLASCYLKQGY	395
DB	1260	LAKQFE--LQLELSKADQESQLODFTSL-KGRHSENGDLVR-----QL	1303
QY	396	AEATLYKELTRAHVOEFGSVDDDDHKPTIMHAEEREEM---SKSRHHEG---GTPYAE	448
DB	1304	EDAESQVQ--LTRLKSLQTSLEEARTRADEARERQTVAAQAKNYQHEAEQLESLEE	1362
QY	449	YGGWYKACKVSSPTVNTLNLGALYRRQKLEAAETLEECALRSRQGTDPISQTKVAE	508
DB	1363	IEKNILKQLS-KANADIQWKARFEGEGLKADE-LED---AKRQQAQKINE--DQE	1414
QY	509	LLGESDGRTSQEGP-----GDSVKEFGGEDASVAVE-WSGDGSGTQLQSGSLGKIRDVL	562
DB	1415	ALDAANSKNASLEKTKSRVLGDL-----DDAQVDVERANGVASALEKKQKGFIDW	1468
QY	563	RRSELLVRLKQGTPE--RPSSNNMKRAASLNYLNQPSAAPLOYRGLSASTMDLS	616
DB	1469	RKATDILAAELDGAQDRDLNTSTDLFAKNA-----QEELA--EVVEGURRENKSLS	1518

RESULT 9  
 IF3X\_DICDI  
 ID IF3X\_DICDI STANDARD; PRT; 1321 AA.

AC O15818;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 15-FEB-2000 (Rel. 39, Last annotation update)  
DE PUTATIVE EUKARYOTIC TRANSLATION INITIATION FACTOR 3 SUBUNIT (EIF-3).  
GN CLUA.  
OS Dictyostelium discoideum (Slime mold).  
OC Eukaryota; Dictyostellida; Dictyostelium.  
RP [1]  
RX SEQUENCE FROM N.A.  
RX MEDLINE; 97352794.  
RA Zhu Q., Hulen D., Liu T., Clarke M.;  
RT "The clua-mutant of Dictyostelium identifies a novel class of  
RL proteins required for dispersion of mitochondria."  
RL Proc. Natl. Acad. Sci. U.S.A. 94:7308-7313(1997).  
CC -|- FUNCTION: BINDS TO THE 40S RIBOSOME AND PROMOTES THE BINDING OF  
CC METHIONYL-TRNAI AND MRNA (BY SIMILARITY).  
CC -|- FUNCTION: REQUIRED FOR PROPER DISPERSION OF MITOCHONDRIA.  
CC -|- SUBUNIT: EIF-3 IS COMPOSED OF AT LEAST 10 DIFFERENT SUBUNITS (BY  
CC SIMILARITY).  
CC -|- SIMILARITY: BELONGS TO THE EIF3 P135 FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; U49332; AAC26527.1; .  
CC DICTYDB; D007???; CLUA.  
KW Initiation factor; Protein biosynthesis.  
FT DOMAIN 178 185 POLY-GLN.  
FT DOMAIN 562 565 POLY-GLU.  
FT DOMAIN 1250 1263 POLY-THR.  
FT SEQUENCE 1321 AA; 148917 MW; 94B2E1F83159437E CRC64;  
SQ  
  
Query Match 5.6%; Score 178.5; DB 1; Length 1321;  
Best Local Similarity 21.2%; Pred. No. 0.0086;  
Matches 108; Conservative 79; Mismatches 216; Indels 107; Gaps 17;  
  
QY 129 RLQRSEQAVALLEEKKHLE-FLGLRQVDEGHTSEKEDATKSDLDLFPNEEEDP 187  
DB 826 RLRLSTNASDMAHSISHFLNCLFLG-----TETGVSADKSKKAK-----QIKSSA 871  
QY 188 SNGLSRGQ-----GATAAQGGYEIP-----ARLRLHNLVIOYAAQ-----GRYEVA 230  
DB 872 INELTQGLWSEIAQLVSKSFDFEIPTHSVPMESRLVLRICLKMGIQILAKDYNFTD 931  
QY 231 VPLCKQALDELRTSGRHPDVATMLNII--ALVYRDQNKYKEAAHLLNDALSIRESTLG 288  
DB 932 APSPEIDVDFPIVXHVNPSTDGLDLEAGKTFNQRYKQYELATELLGEALAIYHVHG 991  
QY 289 PDHPAATAUNNAVLVYKRGYKEAPICQORALETRKVLGNTNPDVAKQLNNALLCQ 348  
DB 992 PIHPDAGACFTHLALAYQNEQYDLAIEYQKNALVITEKTGLDHHETVQAYTTLAVFCQ 1051  
QY 349 NQCKYEAVERVYORALAIYEGQLGPNPNVARTKKNLASCYLKQKYABEATLYKEILTR 408  
DB 1052 RSGRYNESIGYKXVLYLTDLGGEYNPERASYITAAILEDTERFDLAEFLKQLKH 1111  
QY 409 AHVQEGSVDDDDHKPIIMHAEEREEMSKSRHHHEGGTPYAEYGGWYKACKVSSFTVNTILR 468  
DB 1112 ---QEFLETPD-----HLMC-----STT-- 1126  
QY 469 NLGALYRQCKLEAAETLEBCLARSROGTDPISTQKVAELLEGSDGRTSQEGPDSVK 528  
DB 1127 -----YHKMAIVCARATNFDDSIHOKKSTDIILEK-----LGEAH-PTKQ-----SLE 1170  
QY 529 FEGGEDASVAVESGGSGTQLORSGLGKTRDYLRRSSSELLVVRKLGQTEPR-----PSSS 583

DB 1171 FYTG-----LSQTANQIKLFKQHQALKAEODELARLOKEKADQFKKSQPRVSNMPPSLE 1224  
QY 584 NMKRAASLNLYLN-OPSAAPLQVSRGLSAST 612  
DB 1225 NGSVSELLNYNGPKKSQSKSKSTNTTT 1254

RESULT 10  
SR72\_CANFA  
ID SR72\_CANFA STANDARD; PRT; 670 AA.  
AC P33731.  
DT 01-FEB-1994 (Rel. 28, Created)  
DT 01-JUN-1994 (Rel. 29, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE SIGNAL RECOGNITION PARTICLE 72 KDA PROTEIN (SRP72).  
GN SRP72.  
OS Canis familiaris (Dog).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.  
RN [1]  
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
RX MEDLINE; 93273803.  
RX Luetcke H., Prehn S., Ashford A.J., Remus M., Frank R.,  
RA Dobberstein B.;  
RT "Assembly of the 68- and 72-kD proteins of signal recognition  
RT particle with 7S RNA."  
RL J. Cell Biol. 121:977-985(1993).  
CC -|- FUNCTION: SIGNAL-RECOGNITION PARTICLE ASSEMBLY HAS A CRUCIAL ROLE  
CC IN TARGETING SECRETORY PROTEINS TO THE ROUGH ENDOPLASMIC  
CC RETICULUM MEMBRANE. SRP72 BINDS THE 7S RNA ONLY IN PRESENCE OF  
CC SRP68. THIS RIBONUCLEOPROTEIN COMPLEX MIGHT INTERACT DIRECTLY WITH  
CC THE DOCKING PROTEIN IN THE ER MEMBRANE AND POSSIBLY PARTICIPATE  
CC IN THE DOCKING ARREST FUNCTION.  
CC -|- SUBUNIT: SIGNAL RECOGNITION PARTICLE CONSISTS OF A 7S RNA MOLECULE  
CC OF 300 NUCLEOTIDES AND SIX PROTEIN SUBUNITS: SRP72, SRP68, SRP54,  
CC SRP19, SRP14 AND SRP9.  
CC -|- SUBCELLULAR LOCATION: CYTOPLASMIC.  
CC -|- DOMAIN: THE C-TERMINUS IS ESSENTIAL FOR THE INTERACTION WITH THE  
CC SRP68//S RNA COMPLEX.  
CC -|- SIMILARITY: BELONGS TO THE SRP72 FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; X67813; CAA48014.1; .  
CC PIR; S32167; S32167.  
CC PIR; A40692; A40692.  
KW Signal recognition particle; Ribonucleoprotein.  
FT INIT\_MET 0 0 PROBABLE.  
FT MOD\_RES 1 1 BLOCKED.  
FT DOMAIN 551 560 POLY-LYS.  
FT DOMAIN 661 664 POLY-LYS.  
FT SEQUENCE 670 AA; 74362 MW; D394CC56600B5C3D CRC64;  
SQ  
  
Query Match 5.6%; Score 175.5; DB 1; Length 670;  
Best Local Similarity 21.1%; Pred. No. 0.0055;  
Matches 129; Conservative 88; Mismatches 234; Indels 161; Gaps 28;  
  
QY 116 NQWLRELAGTQQLQRSEQAVALLE---EKKHLEFLGQ----LRQYDEGHTSEEK 167  
DB 76 NNSLSFKAYCEYKRLNRIENALKTIESANQOTDKLELYQVLYRLRYDE----CLAVY 131  
QY 168 GDATAKSDLDLFPNEEEDPSNGLSRGGGATAAQGGYE--IPARL-----RTLHNLV 218  
DB 132 RDLVRSQDDY-----DEERTNLS-----AVVAAQSNWKEVWPNENGLQEGTHELCYNAA 182  
QY 219 IQYAAQGRYEVAV-----PLCKQAL-EDLERTSGRHPDVATMLNIALVYRDQNKYK 270



Q01202; p90711;  
 AC 01-FEB-1996 (Rel. 33, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 15-JUL-1998 (Rel. 36, Last annotation update)  
 DE PARAMYOSIN.  
 OS Brugia malayi.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;  
 OC Onchocercidae; Brugia.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Langy S., Luquaud P., Nicolas L.;  
 RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE OF 180-263 AND 684-880 FROM N.A.  
 RX MEDLINE; 92131068  
 LI B., Chandrasekar R., Alvarez R.M., Liftis F., Weil G.J.;  
 RT "Identification of paramyosin as a potential protective antigen  
 RT - against Brugia malayi infection in jirds.";  
 RL Mol. Biochem. Parasitol. 49:315-324(1991).  
 CC -1- FUNCTION: PARAMYOSIN IS A MAJOR STRUCTURAL COMPONENT OF  
 CC MANY THICK FILAMENTS ISOLATED FROM INVERTEBRATE MUSCLES.  
 CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: THICK FILAMENTS OF THE MYOFIBRILS.  
 CC -1- SIMILARITY: HIGH, TO MYOSIN HEAVY CHAINS.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; U77590; AAC18613.1; -;  
 CC EMBL; M63097; AAA27859.1; -;  
 CC EMBL; M63098; AAA27860.1; -;  
 CC HSP; P80220; IDIP.  
 CC PFAM; PF01576; Myosin\_tail; 1.  
 CC Coiled coil; Heptad repeat pattern; Muscle protein; Thick filament;  
 CC Myosin.  
 FT DOMAIN 1 34 NONHELICAL REGION (POTENTIAL).  
 FT DOMAIN 35 859 COILED COIL (POTENTIAL).  
 FT DOMAIN 860 880 NONHELICAL REGION (POTENTIAL).  
 FT CONFLICT 189 189 V -> I (IN REF. 2).  
 FT CONFLICT 197 197 D -> A (IN REF. 2).  
 FT CONFLICT 251 251 L -> F (IN REF. 2).  
 FT CONFLICT 255 255 E -> Q (IN REF. 2).  
 FT CONFLICT 596 596 A -> R (IN REF. 2).  
 FT CONFLICT 730 736 FRKLRR -> IQAEETAA (IN REF. 2).  
 FT CONFLICT 825 825 L -> LQ (IN REF. 2).  
 FT CONFLICT 857 880 HQLRAKMLQKQKFTFSKMSNRDN -> SSVVTGKNASAK  
 FT IYVLEDDQ (IN REF. 2).  
 FT SEQUENCE 880 AA; 101904 MW; F79A275642D4A85E CRC64;  
 Query Match 5.48; Score 171.5; DB 1; Length 880;  
 Best Local Similarity 21.88; Pred. No. 0.013;  
 Matches 137; Conservative 110; Mismatches 242; Indels 139; Gaps 29;  
 QY 34 EALRSEHQVLSQTEICQOGHEGLVHEKARQLRRSMENIELGLSEAOVMLALAS 93  
 DB 152 EKRQLQHEVI--ELPATIDQLKKKH-----LAERAAE-RFEAQVEL-----SNKVDLNR 200  
 QY 94 HLSTVESEKQLRAQVRRLCOENOWRLDELAGTQ----QRLQRSEQAVALLEE-EKKHLE 148  
 DB 201 HVNDLAQQRQLQAEANDLLKEIHDKVQLDNLQVQLQAQLEARRRLEDAERSQ 260  
 QY 149 FLGQLROYDEGHTSEEREGATKDSLDLPNEEEDPSNGLSGGQATAAQGGYEIP 208  
 DB 261 LQAQLHQV-----QLEDSVRLTALDESAARAEEHKLALAN----TEITQWKSFD 308  
 QY 209 ARRLRLHNLVIO-----YAAQGRYEVAVPLCKQALEDLERTSGRHGPDVATML----- 256

DB 309 AEV-ALHHEVEEDLRKKMLQKQAEYEEQIIMLQKISLEKAKSRLOSEVEVLIVLEKA 367  
 QY 257 -NILAVYRDQNKYKEAAHLN---DALSIRESTLGPDPAPVAATLNNLAVLYGKRGKY 312  
 DB 368 QNTAILERAKEQLEKTVNELKVRIDELTVELEAAQREARAALAEQKLNLYEKAQEQ 427  
 QY 313 EA-----EPCQCRALREIKREKVLGTNHPDVAQKLNAL-----LCQNGKYEAVE 357  
 DB 428 EALARENKQLQDDLHEAKEALADANRKLHELDLENARLAGEIRELOTALKESEARRDAE 487  
 QY 358 RYQORALAIYEGQLGPDNPVNARTKNLNASCYLKQKGY-----AEAEYKELTR 408  
 DB 488 NRAQRLAAELQ-QRIEMERRLQKEEEMALRKNQWFEIDRLTAALADAEARKAEISR 546  
 QY 409 -----AHVQEFQ-SVDDHDKPIWMAHEEREEMSKSRHHEGGTPYAEYGGWKACKVSSP 461  
 DB 547 LKKYQAEIAELEMTVDNLNR---ANIEAQKTIKQS-----EQLKILQA 588  
 QY 462 TVNTTLRLNG-----ALYRQKLEAAETLEEC-----ALSRROG-----T 498  
 DB 589 SLEDTQRLQOQLDQYALQAKVSAESAE-LEECKVALDNARAKQAEIDLEANGRIT 647  
 QY 499 DPIS-----QTKVAELLEGSDGRRTSQEGPDSVKFEGGEDASVAVVSGSDGS 546  
 DB 648 DLVSNNNLTAIKNLEFELSTAQALD-EATKELHAADERANPALADAAARVDELHE-- 704  
 QY 547 GTLQSGSLGKIRVDVLRSSSELLVRKQLQ 574  
 DB 705 ---EQEHSM-KI-DALRKSLEEQVKQLQ 727  
 RESULT 13  
 MYST\_HUMAN  
 ID MYST\_HUMAN STANDARD; PRT; 1857 AA.  
 AC P35749; O00396;  
 DT 01-JUN-1994 (Rel. 29, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE MYOSIN HEAVY CHAIN, SMOOTH MUSCLE ISOFORM (SMHC) (FRAGMENT).  
 GN MYH11.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Adams M.D., Loftus B.J., Zhou L., Phillips C., Brandon R.,  
 RA Fuhrmann J., Kim U.J., Kerlavage A.R., Venter J.C.;  
 RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE OF 772-1857 FROM N.A.  
 RX MEDLINE; 93263189.  
 RA Matsuo R., Yoshida M.C., Furutani Y., Imamura S., Kanda N.,  
 RA Yanagisawa M., Masaki T., Takao A.;  
 RT "Human smooth muscle myosin heavy chain gene mapped to chromosomal  
 RT region 16q12.";  
 RL Am. J. Med. Genet. 46:61-67(1993).  
 RL [3]  
 RN SEQUENCE OF 978-1857 FROM N.A.  
 RP TISSUE=HIPPOCAMPUS;  
 RC Okajima K.;  
 RL Submitted (NOV-1992) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: MUSCLE CONTRACTION.  
 CC -1- SUBUNIT: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2  
 CC HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC)  
 CC AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).  
 CC -1- SUBCELLULAR LOCATION: THICK FILAMENTS OF THE MYOFIBRILS.  
 CC -1- TISSUE SPECIFICITY: SMOOTH MUSCLE; EXPRESSED IN THE UMBILICAL  
 CC ARTERY, BLADDER, OESOPHAGUS AND TRACHEA.  
 CC -1- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING  
 CC CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES.  
 CC CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.  
 CC -1- PTM: TWO CYSTEINE RESIDUES IN THE S1 DOMAIN ARE SELECTIVELY  
 CC ALKYLATED AND ARE REQUIRED FOR MYOSIN ATPASE ACTIVITY.



```
FT DOMAIN      861      879      NONHELICAL REGION (POTENTIAL).
SQ SEQUENCE    879 AA: 101125 MW; DCA24AC01A9F02B1 CRC64;

Query Match
Best Local Similarity 5.4%; Score 169.5; DB 1; Length 879;
Matches 138; Conservative 107; Mismatches 244; Indels 139; Gaps 29;

QY 34 EALRSEHQVQLSQTTECLQQGHEGLVHKARQLRRSMENIELGLSPAQWMLALAS 93
DB 152 ERQRUQHEVI--ELTAADIDQLQDKH-----LAEKAAE-RFEAQTIEL-----SNKVEDLNR 200
QY 94 HLSTVESKQKLRAQVRLCOENWLRLDELQGTQ----QLRQSEQAVALLEE-EKKHLE 148
DB 201 HVNDLAQORQLAENNDLLHEIDQKVLQDNLNQHVKYQLAQLEEARRRLEDRERSQ 260
QY 149 FLGQURQYDEGHTSEKEGSDATKSDLDLFPNNEEEDPSNLSRGOGATAAQQGGYEIP 208
DB 261 LQAQLHQV-----QLELDSVRLTALDEESARAARAEHKLALAN----TEITQWKSFD 308
QY 209 ALRLTLHLNLIQ-----YAAQGYEVAVPLCKQALDELRTSGRHPDVMML-----256
DB 309 AEV-ALHHEEVEDURKMLQKQAEYEQIETMLQKISQLEKAKSRLOSEVEVLIVDLEKA 367
QY 257 -NILALVYRDONKYKEAAHLN---DALSIRESITLGPDPHPAATLNNLAVLYGKRGYK 312
DB 368 QNTAILERAKEQLEKTVNELKVRIDELTVELEAAQREARAALAELOKMNLYEKAVEQK 427
QY 313 EA-----EPLCORALEIREKVLGNHPDPAQVAKLNNLAL-----LCQNGQKYEAVE 357
DB 428 EALARENKQLQDLHDEAKEALADANRKLHELDENARLAGEIRELQATLAKESAEARDAE 487
QY 358 RYORALAIYEGOLGPDNPVARTKNNLASCYLKQGY-----AEATLYKEILLR 408
DB 488 NRAORALAELO-QRIEMERLQKEBEMEALRKNMDFEIDRLTAALADAEARKAEIAR 546
QY 409 -----AHVQFEG-SVDDHDKPIWNHAEEREMSKSRHHEGGTPYAEYGGWYKACKVSSP 461
DB 547 LKKYQAEIAELEMTVDNLR---ANIEAQTKIKQS-----EQLKVLOA 588
QY 462 TVNTTLNLG-----ALYRQGLKLEAAETLEEC-----ALRSRQG-----T 498
DB 589 SLEDTQROLQOQLDQYALQAKVSAESAE-LEECKVALDNARAKQAEDLEBEANGRIT 647
QY 499 DPIS-----OTKVAELLGSDGRRTSQEGPGDSVKFEGGEDASVAVEWSGDGS 546
DB 648 DLVINNNLTAIKNKLETEUSTAQADLD-EATKELHAADERANRALADAAARAVEQLHE-- 704
QY 547 GTLQSGSLGKIRDVLRSSSELLVKRLQ 574
DB 705 ---EQEHSM-KI-DALRKSLEEYKQLQ 727
```

## RESULT 15

```
ID MYST_RABIT STANDARD; PRT; 1972 AA.
AC P35748;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DE 15-DEC-1998 (Rel. 37, Last annotation update)
DE MYOSIN HEAVY CHAIN, SMOOTH MUSCLE ISOFORM (SMHHC).
GN MYH11.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 92073350.
RA Babij P., Kelly C., Periasamy M.;
RT "Characterization of a mammalian smooth muscle myosin heavy-chain
RT gene: complete nucleotide and protein coding sequence and analysis of
RT the 5' end of the gene.";
RL Proc. Natl. Acad. Sci. U.S.A. 88:10676-10680(1991).
```

```
CC -!- FUNCTION: MUSCLE CONTRACTION.
CC -!- SUBUNIT: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2
CC HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC)
CC AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
CC -!- SUBCELLULAR LOCATION: THICK FILAMENTS OF THE MYOFIBRILS.
CC -!- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
CC CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
CC CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
CC -!- PTM: TWO CYSTEINE RESIDUES IN THE S1 DOMAIN ARE SELECTIVELY
CC ALKYLATED AND ARE REQUIRED FOR MYOSIN ATPASE ACTIVITY.
CC -!- MISCELLANEOUS: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT
CC MEROMYOSIN (LMN) AND 1 HEAVY MEROMYOSIN (HMM). IT CAN LATER BE
CC SPLIT FURTHER INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED
CC SUBFRAGMENT (S2).
CC -!- SIMILARITY: BELONGS TO THE MYOSIN HEAVY CHAIN FAMILY.
CC
CC THIS SWISS-PROT ENTRY IS COPYRIGHT. IT IS PRODUCED THROUGH A COLLABORATION
CC BETWEEN THE SWISS INSTITUTE OF BIOINFORMATICS AND THE EMBL OUTSTATION -
CC THE EUROPEAN BIOINFORMATICS INSTITUTE. THERE ARE NO RESTRICTIONS ON ITS
CC USE BY NON-PROFIT INSTITUTIONS AS LONG AS ITS CONTENT IS IN NO WAY
CC MODIFIED AND THIS STATEMENT IS NOT REMOVED. USAGE BY AND FOR COMMERCIAL
CC ENTITIES REQUIRES A LICENSE AGREEMENT (SEE HTTP://WWW.ISB-SIB.CH/ANNOUNCE/
CC OR SEND AN EMAIL TO LICENSE@ISB-SIB.CH).
CC
CC EMBL; M77812; AAA31395.1; -.
CC PIR; A41604; A41604.
CC HSP; P08799; 1MND.
CC PFAM; PF00612; IQ; 1.
CC PFAM; PF01576; Myosin_tail; 1.
CC PFAM; PF00063; Myosin_head; 1.
CC PRINTS; PR00193; MYOSINHEAVY.
CC
CC MYOSIN: Muscle protein; Coiled coil; Thick filament; Actin-binding;
CC ATP-binding; Methylation; Alkylation; Multigene family.
CC
CC DOMAIN 1 843 RODLIKE TAIL (S2 AND LMN DOMAINS).
CC FT DOMAIN 844 1934 CARBOXYL-TERMINAL.
CC FT DOMAIN 1935 1972 ATP (POTENTIAL).
CC FT NP_BIND 178 185 ACTIN-BINDING (BY SIMILARITY).
CC FT DOMAIN 561 583 ACTIN-BINDING (BY SIMILARITY).
CC FT DOMAIN 763 777 IQ.
CC FT DOMAIN 785 807 COILED COIL (POTENTIAL).
CC FT DOMAIN 844 1934 METHYLATION (TRI-) (POTENTIAL).
CC FT MOD_RES 129 129 ALKYLATION (SH-1) (POTENTIAL).
CC FT MOD_RES 701 701 ALKYLATION (SH-2) (POTENTIAL).
CC FT MOD_RES 711 711
CC SEQUENCE 1972 AA; 227318 MW; 2061A224288D6A4C CRC64;
```

```
Query Match
Best Local Similarity 5.3%; Score 166; DB 1; Length 1972;
Matches 136; Conservative 109; Mismatches 265; Indels 136; Gaps 24;

QY 21 EILGSTRVLSQGLRALRSEHQVQLSQTTECLQQGHEGLVHKARQLRRSMENIEL 80
DB 1235 DLAGELVLGQAKQEVHEHKKKLEVLQELQSKSCDGERARAEALNDVKHLQNEVESVTG 1294
QY 81 GLSEAAQ---VMLA-----LASHLSTV-----ESEKQLR--AQVRLCOENQWLRLDELQ 125
DB 1295 MLSEAEGRKAIKLAKEVASLGSQLODTQELLQEETRQKLNVTSTKLRLQEDERNLSQLEQ 1354
QY 136 TQOQLRQSEQA---QLEEEKHLEFLGOLRQYDEGHTSEKEGDAYKSDLDLFPN 181
DB 1355 EWEAKQNLERHISTINTQLSDSKKLQDFASTVESLEEGKKRFQKE-----IESATQ 1407
QY 182 FEEEDPSNLSRGOGATAAQQGGYEIPARL-RTLHNVIQYAAQGYEVAVPLCKQALQ 240
DB 1408 YEEK-----AAAYDKLEKTKNRLQQLDLDLVLDNQ-----RQLVSN 1445
QY 241 LERTSGRHPDPAVMTNLIALVYRDQNKYKEA-----AHLNDAISIRESTLG 288
DB 1446 LEKKQKFDQLLAPEKNI-SSKYADERDRAEAAREKTRKALSRLAEALEAKEE-LE 1503
QY 289 PDHPAATLNNLAVLYGKRGYKEAEPQLQAL-----EIREKVLGTHNPDV 336
```

Db 1504 RTNKM LKAEMEDLVSSKDDVGNVHLEKSKRALETQMEEMTQLBEELELQATEDAKL 1563  
QY 337 AKOLNNLALCONOGKVEAVE-----RYQALAIYEGQLGPDNPNTVARTKNNLASCY 389  
Db 1564 RLEVNQALKVQFERDLQARDEQNEKRRQLQLQHEIETELED-----RKORALAAAA 1618  
QY 390 LK--QGYAE-----AETLYKEIL-TRAHVQEFSGVDDDHK-----PIWMHAER 431  
Db 1619 KKKLEGDLKLELQADSAIKGEEAIKQLLKQAQMKDFQRELEDAARASRDEIFATAKEN 1678  
QY 432 EEMSKSRHHGGTPYAEYGGWKACKVSSPTVNTTLNLGAL--YRQCKLEAAETLEEC 489  
Db 1679 EKKAKSLEAD-----LMQLOEDLAAAEARAKQADLEKEELAEEL 1717  
QY 490 AL----RSRROGTDPIQTKVAELLGESDGRRTSQSGPGDSVKFEGGEDASVAVEMSGDG 545  
Db 1718 ASSLSGRNALQDEKRLKLEARIAOLEEELBEEGNGMEAMSDRVKATQQAELSNELATER 1777  
QY 546 SGTLQSGSLGKIRDYLRKSELVLRKLGQTGTEPRPSSNNMKRAASL 591  
Db 1778 S-TAQKNES---ARQQLERQNKELKSKLOEMEGAVKSKFKSTIAAL 1819

Search completed: August 15, 2000, 13:20:24  
Job time: 8881 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 15, 2000, 10:46:30 ; Search time 188.37 Seconds  
(without alignments)  
227.838 Million cell updates/sec

Title: US-09-036-614A-1

Perfect score: 3161  
Sequence: 1 MSGVLGQDEPAGHRLSQE.....APLQVSRGLSASTMDLSSSS 619

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 225878 seqs, 69334122 residues

Total number of hits satisfying chosen parameters: 225878

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_12.\*

- 1: sp.archaea.\*
- 2: sp.bacteria.\*
- 3: sp.fungi.\*
- 4: sp.human.\*
- 5: sp.invertebrate.\*
- 6: sp.mammal.\*
- 7: sp.mhc.\*
- 8: sp.organella.\*
- 9: sp.phase.\*
- 10: sp.plant.\*
- 11: sp.rodent.\*
- 12: sp.virus.\*
- 13: sp.vertebrate.\*
- 14: sp.unclassified.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2056.5	65.1	599	11	O88448 mus musculu
2	1910	60.4	541	11	O88447 mus musculu
3	1775	56.2	487	13	Q90707 gallus gall
4	962.5	30.4	521	5	Q21592 caenorhabdi
5	503.5	15.9	490	2	Q33965 plectonema
6	276.5	8.7	1049	2	Q9XB26 myxococcus
7	232	7.3	609	10	O81629 arabidopsis
8	229.5	7.3	886	2	O86768 streptomyce
9	196.5	6.2	707	10	Q9ZW40 arabidopsis
10	189	6.0	342	2	O67735 aquifex ae
11	189	6.0	948	2	Q55489 synecocyst
12	181	5.7	1939	5	Q25662 plasmodium
13	179.5	5.7	2176	5	O46112 drosophila
14	179	5.7	1963	5	O02244 caenorhabdi
15	178	5.6	1011	5	Q9Y148 drosophila
16	176.5	5.6	1604	10	Q9X140 arabidopsis
17	167.5	5.3	2101	4	Q14981 homo sapien
18	167.5	5.3	3187	11	Q63714 rattus norv
19	166	5.3	914	10	Q96301 arabidopsis

20	166	5.3	1938	11	O08639
21	166	5.3	1972	11	O08638
22	164	5.2	2442	4	O60588
23	163	5.2	1313	4	O75033
24	163	5.2	2168	11	O61043
25	162	5.1	1992	5	Q21440
26	162	5.1	2442	4	O14812
27	161.5	5.1	1133	5	O21022
28	161.5	5.1	2115	4	O14980
29	160.5	5.1	944	10	O82422
30	160.5	5.1	1426	4	O60314
31	160.5	5.1	1756	4	O60454
32	160	5.0	1134	4	O75153
33	159.5	5.0	543	11	O54981
34	159	5.0	1974	5	O21000
35	158.5	5.0	3321	4	O95613
36	158	5.0	1743	5	O96063
37	158	5.0	1957	5	O04010
38	157.5	5.0	1941	5	O26079
39	157.5	5.0	2230	4	O13439
40	157	5.0	886	1	O29230
41	157	5.0	1257	4	O75334
42	157	5.0	1935	5	O44934
43	156.5	5.0	1390	5	O77033
44	156.5	5.0	1958	5	O96062
45	156.5	5.0	1999	11	O63731

#### ALIGNMENTS

RESULT 1

O88448  
ID O88448 PRELIMINARY; PRT; 599 AA.  
AC O88448;  
DT 01-NOV-1998 (TREMBlrel. 08, Created)  
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)  
DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)  
DE KINESIN LIGHT CHAIN 2.  
GN KLC2.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BALB C; TISSUE-BRAIN;  
RX MEDLINE; 96288268.  
RA RAHMAN A., FRIEDMAN D.S., GOLDSTEIN L.S.;  
RT "Two kinesin light chain genes in mice. Identification and  
RT characterization of the encoded proteins.";  
RL J. Biol. Chem. 273:15395-15403(1998).  
DR EMBL; AF055666; AAC27741.1.  
DR PFAM; PF00409; kinesin2.4.  
DR PRINTS; PR00381; KINESINLIGHT.  
SQ SEQUENCE 599 AA; 66662 MW; 0B012E26 CRC32;

Query Match 65.1%; Score 2056.5; DB 11; Length 599;  
Best Local Similarity 69.6%; Pred. No. 1.3e-119;  
Matches 422; Conservative 61; Mismatches 102; Indels 21; Gaps 8;

QY 1 MSGVLGQDEPAGHRLSQEILGSLVSGLEALRSEHQAVLQSLQSQTIECQQGHE 60  
Dd 1 MATWV-----PREEKLSQDEIVLTKAVIQGLTGRGHEHALLAPLAS-----HEAGEA 50  
QY 61 EGLVIEKARQRRSMENTELGLSEAQVLMALASHLSTVESEKQKLRQVRLQENQWLR 120  
Dd 51 EPGSQERCLLRSLLEATELGLGEAQVILASSHLGAVESEKQKLRQVRLQENQWLR 110  
QY 121 DELACTQQLORSEQAQVLALEEKHKLFLQLQRYDEDDGTSEEKGDATKDSLDLFF 180  
Dd 111 DELACTQQLORSEQAQVLALEEKHKLFLQRYDEDDGTSEEKGDVFKDSLDDLFF 168



```
Db 216 LNILALVYRDQNKYKDAANLNDALAIAREKTLGKDHDPAAVATLNNLAVLYGKRGYKEAE 275
QY 316 PLCQRALEIREKVLGTNHPDPAKQNLNLLALCQNGKYEAVERYYQALAIYEGOLGPDN 375
Db 276 PLCKRALEIREKVLGKDPDAKQNLNLLALCQNGKYEEVEYYQALAIYQKLGPD 335
QY 376 PNVAKTNNLASCYLKQKGYAEATLYKEILTRAHVQFSGVDDHKKPIWMHAEREEMS 435
Db 336 PNVAKTNNLASCYLKQKGYAEATLYKEILTRAHERFSGVDENKPIWMHAERECK 395
QY 436 KSRHHEGTPVAEYGGWKYKACKVSPVTNTLNLGALYRQGLERAETLEECALSR 495
Db 396 GKQ--KQISFGEYGGWKYKACKVSPVTNTLNLGALYRQGLERAETLEEAAMRSR 453
QY 496 QGTDPISTQVAELIGESDGRRTSQ 520
Db 454 QGLDNVHKQVAEVLNDPESIEKSE 478

RESULT 4
ID Q21592 PRELIMINARY; PRT; 521 AA.
AC Q21592;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-NOV-1999 (Tremblrel. 12, Last annotation update)
DE M7.2 PROTEIN.
GN M7.2.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
OC Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
FN [1]
RA SEQUENCE FROM N.A.
RA LENARD N.
RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RN SEQUENCE FROM N.A.
RX MEDLINE; 94150718.
RA WILSON R., AINSOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA CRAXTON M., DEAR S., DU Z., DUREIN R., FAVELLO A., FULTON L.,
RA GARDNER A., GREEN P., HANKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIFEK N., ROOPRA A., SAUNDERS D., SHOWNKEN R.,
RA SWALDON N., SMITH A., SONNHAMMER E., STADEN R., SULTON J.,
RA THERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOLDMAN P.;
RA "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans."
RL Nature 368:32-38(1994).
DR EMBL; Z68337; CAA92746.1; -.
DR PFAM; PF00409; kinesin2; 4.
DR PRINTS; PR00381; KINESINLIGHT.
SQ SEQUENCE 521 AA; 58649 MW; 1BD379DB CRC32;
```

Query Match 30.4%; Score 962.5; DB 5; Length 521;  
Best Local Similarity 44.0%; Pred. No. 4.5e-52;  
Matches 220; Conservative 83; Mismatches 166; Indels 31; Gaps 11;

```
QY 20 EEILGSTRVLSQGLAEALRSEHQA---VLOS--LSOTIECIQQGHEGLVHEKAROLRRS 74
Db 7 DDITVLTQVQTLFALRDEHEATRLLEANLISKIYIGNSS--EFSUPSEKMGUIDS 64
QY 75 MENTELGSEAQVMAALASHLSTVSESKQLRAQVRRLCOENQWLDELQAGTQORLQ 134
Db 65 LGKVMGDGDEASLLIMDKLMQSYDVQLSKNHESIRLLRQENTWLLDELTTIQKIQESE 124
QY 135 QAVAQLEEEKKHLEFLGQLQRYDED--GHTSEKEGDAKDKSLDDLPNPEEEDPSNGLR 193
Db 125 RTVAHLEERDHYKFDQSMNLYNSDFQHTTSVDATPMVMDTLQELGFGPEEDQNNN-QA 183
```

```
QY 194 GOGATAAQ-----QGGYEIPARLRTLHNLVIOYAAQGRYEAVALPCKQALDLERTSGR 248
Db 184 DQCRSSSFNPNISNDYQPLRLQTLQNLVIOYMEQGRFEVAIPCKQALDLEVVHGVN 243
QY 249 HPDVATMLNLALVYRDQNKYKEAAHLLNDALSIRESTLGPDPHPAATAATLNNLAVLYGKR 308
Db 244 HLDVATMLNLVLAIVYRNQENFKDAIYLEKALSIRVQCCGENHSHVAATLNLATAYGR 303
QY 309 GYKAEAPLCQRALEIREKVLGTNHPDPAKQNLNLLALCQNGKYEAVERYYQALAIYE 368
Db 304 GYKSEPLECKRALEIRKKNLGNHPDPAKQNLNLGIVTQOLEKYETEENYFKQALSYN 363
QY 369 GOLGPDNPVARTKNNLASCYLKQKGYAEATLYKEILTRAHVQFSGVDDHKKPIWHA 428
Db 364 RAPPHENQNVIKTKQLASVFLKQKGYAEEMYNILSKVAI-----TGKPKPIWIA 416
QY 429 EREEMSKSRHHEGTPYAEYGGW-YKACKVSSPTVNTLNLGALYRQGLERAETLE 487
Db 417 EDREERQRN----GIPKVDDESEFNVTVDNSVMSTIKNLAAYVRKQKEAAGTLE 471
QY 488 ECALRSRRO---GTDPISTQ 504
Db 472 E-ALGAKQINGADHTNST 490

RESULT 5
ID Q33965 PRELIMINARY; PRT; 490 AA.
AC Q33965;
DT 01-JAN-1998 (Tremblrel. 05, Created)
DT 01-JAN-1998 (Tremblrel. 05, Last sequence update)
DT 01-NOV-1998 (Tremblrel. 08, Last annotation update)
DE KINESIN LIGHT CHAIN.
GN KLC.
OS Plectonema boryanum.
OC Bacteria; Cyanobacteria; Oscillatoriales; Plectonema.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-BTEX 485;
RX MEDLINE; 97355691.
RA CELERIN M., GILPIN A.A., DOSSANTOS G., LAUDENBACH D.E., CLARKE M.W.,
RA BEUSHAUSEN S.;
RT "Kinesin light chain in a eubacterium."
RL DNA Cell Biol. 16:787-795(1997).
DR EMBL; U78597; AAB87735.1; -.
DR PFAM; PF00409; kinesin2; 5.
SQ SEQUENCE 490 AA; 55967 MW; 4A7472E2 CRC32;
```

Query Match 15.9%; Score 503.5; DB 2; Length 490;  
Best Local Similarity 33.9%; Pred. No. 9.7e-24;  
Matches 164; Conservative 62; Mismatches 153; Indels 105; Gaps 17;

```
QY 130 LORSEQAQVAOLEEEKKHLEFLGQLRQYDEDEGHTSEK-----EGDATKSLDLPNE 182
Db 73 LKRVQGMQLQLLUR--EFFVVKREQRADSDQLQRFYQVIVAEAKVRD----- 121
QY 183 EEPEDNSLGRQOGATAAQ--QGGYEIPAR-----LRTLHLNV--IOYAAQGRYEAVALP 233
Db 122 ---EPEKSLIRESTMIAHLQEMERLARPEQALDLATCLNWLAEIYVQGRYEEAEPL 178
QY 234 KQALDLEERTSGRHPDVATMLNLALVYRDQNKYKEAAHLLNDALSIRESTLGPDPHPA 293
Db 179 VYRSLSIHQQLGADHLDVANSFNLLALYKEQGRYEEAEPLVYRSLRSIRSQQLGTDHLD 238
QY 294 VAATLNNLAVLYGKRGYKEAPLCQRALEIREKVLGTNHPDPAKQNLNLLALCQNGKY 353
Db 239 VATSLNNLAVLYGSGRYEEAEPLVYRSLRSIREQQLGTDHLDVATSLSNLAVLYSQSGCH 298
QY 354 EAVERYQALAIYEGOLGPDNPVARTKNNLASCYLKQKGYAEATLYKEILTRAHVQ 413
Db 299 HKAEPILVRAIPVWEQQLGTDHPDVATSLNLAFLYHLQGRYEEAEPLVRAALS--IREQ 357
```



Db 315 NIYMSLCRDEAVSFYKOKALTVEKASKGTHPTVASVFVRLAELHRTGKLRSEKSYCN 374  
 QY 401 ---LYKEIITRAHVQFGVDDDDHPIWMAHEEREMSKSRHEGGTYPAYEGGWYKACK 457  
 Db 375 ALRIYNKPVPGTVEEIIAGGLTEISAIVSVDEPEALKLQ-----KSMK 420  
 QY 458 V--SSPTVNTLRLN---GALYRROGKLEAAETLEECALRSRROGTD-----ISQT 504  
 Db 421 LLEDKPGQOASAGLARMGVMTYVGRYEDARNAFSAVTKLRAAGEKSAFFGVVLNQ 480  
 QY 505 KVA-----ELLGSDGRRTSQEGFGDSVKFEGGEDASVAVESGDSGFTLRSQ 553  
 Db 481 GLACVOLFRIKIDAGELFEARGILQERQPCDQ-----DLGVYSNLAAYD 527  
 QY 554 SLGKIRD---VLRSESLVRKLOGTPE-----RPSSSNKRAASLNQ 596  
 Db 528 AMGRIDEAIEILQVQLKREELGTANPDQFDEKRLAELLKAGRSRNYRAKSLQNLID 587  
 QY 597 PSAAP 601  
 Db 588 PNARP 592

RESULT 8  
 ID 086768 PRELIMINARY; PRT; 886 AA.  
 AC 086768  
 DT 01-NOV-1998 (Tremblrel. 08, Created)  
 DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)  
 DT 01-NOV-1998 (Tremblrel. 08, Last annotation update)  
 DE PUTATIVE ATP /GTP-BINDING PROTEIN.  
 GN SC6A9.38.  
 OS Streptomyces coelicolor.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomycetes.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-A3(2);  
 RA MURPHY L., HARRIS D.;  
 RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-A3(2);  
 RA PARKHILL J., BARRELL B.G., RAJANDREAM M.A.;  
 RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-A3(2);  
 RX MEDLINE: 97000351.  
 RA REDENBACH M., KIESER H.M., DENAPALTE D., EICHNER A., CULLUM J.,  
 RA KINASHI H., HOPWOOD D.A.;  
 RT "A set of ordered cosmid and a detailed genetic and physical map for  
 the 8 Mb Streptomyces coelicolor A3(2) chromosome";  
 RL Mol. Microbiol. 21:77-96(1996).  
 DR EMBL: AL031035; CAA19923.1; ..  
 SQ SEQUENCE 886 AA; 97715 MW; 848BB121 CRC32;

Query Match 7.3%; Score 229.5; DB 2; Length 886;  
 Best Local Similarity 23.5%; Pred. No. 1.8e-06;  
 Matches 113; Conservative 74; Mismatches 216; Indels 77; Gaps 14;  
 QY 1 MSGLVIGORDEPAGHRLSQEELSTRL-----VSQGLEALRSEHQAVLSQSOTIE 52  
 Db 341 LQGRVGMDEQADTLAAD---LGLDPLALAAQAGVIANGVSLGLYHQLL---ATNTR 394  
 QY 53 CLQQGHEGLVHEKARQLRRSMENIELGLSEAQVNLALASHL-----STVSEKOKL 105  
 Db 395 ILERGQ-APGYAPAPAAVTIAVDRLTDEHPDAIALRLVALLGPEPIRNAWLESVRPL 453  
 QY 106 RAQVRLCQENOWLDELQAGTQRLQSQEVAQVLEEKHLEFLGQ--LR-----154  
 Db 454 -TTIPGSDDDPWLHEALS-----PLSRVGLARTEPFESFQIHLRTQAVLRDQSLPDQSA 506

QY 155 --QYDEGHTSEKEGDKD-----SLDDLFPNEEE-----DPSNGLSR 193  
 Db 507 AIRSDATALLSAANPDGPPQSPGNWPGWAALTPLHTAQHLAPTQPELPTLLDAAHYLR 566  
 QY 194 -GQATA-----AQGGYEIPARLRTLHNLVQYAAOGRYEVAVPCKQALEDL 241  
 Db 567 SGOTRTARDLTATAHAAWATDLDGEDHPTLTLSAQFLGHATADLGEHAEARRIETLRR 626  
 QY 242 ERTSGRHPDVATMLNIALVYRDQNKYEAHLLNDALSIRESTLGPDPHPAATAAFLNL 301  
 Db 627 RRLTGDDHPTLQSANLANVLDNLGEHTESHRIEDTLRRRTTIGDDHPHTLQSAHNL 686  
 QY 302 AVLYGKRGKYKEAPLQCRALEREKVLGTHNPDPVAKQLNNLALLCQNOGKYEAVERYQ 361  
 Db 687 AASLHHLGHEATEARRILEDTLRRRTTIGDDHPTLQSAHLSAASLSLGEHTESYRMDQ 746  
 QY 362 RALAIYEGQLGPDNPVARTKNKLASCYLKQGYAAETLYKEILTRAHVQEGSVDDDH 421  
 Db 747 ETLARLRILGDDHPTLQSAHNLAVTLHDLGHEATEARRIETLRR---QRTLGGDDH 802

RESULT 9  
 ID 092W40 PRELIMINARY; PRT; 707 AA.  
 AC 092W40;  
 DT 01-MAY-1999 (Tremblrel. 10, Created)  
 DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)  
 DT 01-MAY-1999 (Tremblrel. 10, Last annotation update)  
 DE F17L21.30.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;  
 OC core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;  
 OC Arabidopsids.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA SHINN P., BUEHLER E., DEWAR K., FENG J., KIM C., LI Y., SUN H.,  
 RA CONWAY A., CONWAY A., KURTZ D., OJI O., SHEN Y.K., TORIUMI M.,  
 RA VYSOTSKAIA V., YU G., DAVIS R.W., FEDERSPIEL N.A., THEOLOGIS A.,  
 RA ECKER J.R.;  
 RT "Genomic sequence for Arabidopsis thaliana BAC F17L21.";  
 RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AC004557; AAC80630.1; ..  
 SQ SEQUENCE 707 AA; 77828 MW; 89334300 CRC32;

Query Match 6.2%; Score 196.5; DB 10; Length 707;  
 Best Local Similarity 21.6%; Pred. No. 0.00015;  
 Matches 123; Conservative 85; Mismatches 226; Indels 135; Gaps 22;  
 QY 10 DEPAGHRLSQEELT--GSTRLVSQ--LEALRSEHQAVLSQSOTIECLQGG-----58  
 Db 194 EEPSSENVELARLLNQARNLVSSGDSSTHKALELTHRA-----AKLFEASAENGKPCLEW 248  
 QY 59 -----HEEGLVHEKARQLRRSMENIELGLSEAQVNLALASHLSLSTVESEKOKLRAQVRLCQ 114  
 Db 249 IMCLHYTAAVHCKLKE-----YNEAIPVLSORSVEIPVVEGEHALAKFAGLMQ 297  
 QY 115 ENOWLDELQAGTQRLQSQEVAQVLEEE---EKKHL-----EFLGQLRQ 155  
 Db 298 -----LGDITYAMVGQLESSICVTEGLNIQKVLGENDPRVGTCTRYLAELVQALR 349  
 QY 156 YDEGHTSEKEGDKDKSDDLFPNEEEDPSNLSRGOGATAAQGGYEIPARLRTLH 215  
 Db 350 FDEAQVCETA-----LSIHRESGLP-----GSIAPAAD-----RRLM 382  
 QY 216 NLVQYAAOGRYEVAVPCKQALEDELTSGRGHPDVATMLNIALVYRDQNKYEAHNL 275  
 Db 383 GLICE--TKGDHENA--LEHLVLASMAAMAANGOESEVAFVDTISIGDSYLSLSRDEAICA 438  
 QY 276 LNDALSIRESTLGPDPHPAATAAFLNLAVLYGKRGKYKEAEPICORALEIREK-VLGTNHP 334

Db 439 YOKSLTAKTAKGENHPAGSVVIRLADLYNRGKGVREAKSYCNALRIYESHNLISPE 498  
 QY 335 DVAKQNLNALLCONQKGYPAVERYYQALAIYEGOLGPDNP-----VARTNNLASCVL 390  
 Db 499 EIASGLTDISVICSMEVNEVQAITLLQKALKIY-----ADSPQKTMIAIEAQMGVLXY 553  
 QY 391 KQKYAAEAETLYKEILTRAH-----VQFSGVDDHDKPIWHAEREEMSKS 437  
 Db 554 MMGYMESYNTFRKSAISKRLATKCKOSTFFGIALNQMGACIQOLDIAEAEVLFEEAKCI 613  
 QY 438 RHHEGGPYAEYGGWKACKVCKSVPTVNTTLRNGLGYRRQKLEAAETL--BECAALRSR 495  
 Db 614 LEQECGPYHETGLYS-----NLAGAYDAIGRLDDDAIKLLGHVGVVREK 659  
 QY 496 QGT-DPISQ---TKVAELL---GESDGR 517  
 Db 660 LGTANPVTDEKRRLAQLLKEAGNVTRK 688  
 RESULT 10  
 ID O67735 PRELIMINARY; PRT; 342 AA.  
 AC O67735;  
 DT 01-AUG-1998 (T-EMBLrel. 07, Created)  
 DT 01-AUG-1998 (T-EMBLrel. 07, Last sequence update)  
 DT 01-MAY-1999 (T-EMBLrel. 10, Last annotation update)  
 DE HYPOTHETICAL 40.4 KD PROTEIN.  
 GN AQ1896.  
 OS Aquifex aeolicus.  
 OC Bacteria; Aquificales; Aquificaceae; Aquifex.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-VF5;  
 RX MEDLINE; 98196666.  
 RA DECKERT G., WARREN P.V., GAASTERLAND T., YOUNG W.G., LENOX A.L.,  
 RA GRAHAM D.E., OVERBECK R., SNEAD M.A., KELLER M., AUJAY M., HUBER R.,  
 RA FELDMAN R.A., SHORT J.M., OLSON G.J., SWANSON R.V.;  
 RT "The complete genome of the hyperthermophilic bacterium Aquifex  
 aeolicus.";  
 RL Nature 392:353-358(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-VF5;  
 RA DECKERT G., WARREN P.V., GAASTERLAND T., YOUNG W.G., LENOX A.L.,  
 RA GRAHAM D.E., OVERBECK R., SNEAD M.A., KELLER M., AUJAY M., HUBER R.,  
 RA FELDMAN R.A., SHORT J.M., OLSON G.J., SWANSON R.V.;  
 RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AE000762; AAC07708.1; -.  
 DR PFAM; PF00515; TPR; 3.  
 KW Hypothetical protein.  
 SQ SEQUENCE 342 AA; 40375 MW; 58842C97 CRC32;

Query Match 6.0%; Score 189; DB 2; Length 342;  
 Best Local Similarity 25.1%; Pred. No. 0.00016;  
 Matches 88; Conservative 37; Mismatches 119; Indels 106; Gaps 13;  
 QY 221 YAAQGYEAVPLCKQALDELTSGRHPDVATMLNLALVYRDQNKYKEAHLNDAL 280  
 Db 36 FAAR-NYSEALANPHKAL-----RANPDEPRINWALGLAYWEAKYKAESEPKAL 86  
 QY 281 SIRESTLGPDPHAPVAATNLNLAVLYKRGKYKEAPLCQRA----- 321  
 Db 87 SI-----NPNYSEARKNLGILYKLYGEEALYQEAANDYEYKKEAFYVLAKV 138  
 QY 322 -----LEIREKVLGNHPDVAKOLNALLCONQKGYEAVERYORALAIYEGOL 371  
 Db 139 YEAKQDLKNVRYLEKAVANPNFVQAL--ELAQVENLKGVEAEKIKYSLLL----- 191  
 QY 372 GPDNPVARTKNNLASCYLKGQYAE-----ETLYKEILT---RAHVQEFSG----- 416  
 Db 192 --NGFNKPLKYLAEVYKKGQYERAREIIEKLLYKENITNQREKVKELLTKVLLAQ 249

QY 417 ----VDDDHKPIWHAEREEMSKSRHHEGGTPYAEYGGWKACKVSSPTVNTTLRNIGA 472  
 Db 250 RKLIPRVHPI-----KKEKKREKY-----YAV-----OLGA 278  
 QY 473 LYRRQKLEAAETLECAALRSR--QGTDPISQTKVAELLGESDGRITSQE 521  
 Db 279 FSTERADKLVLQELKSLGLRDLRLPTDGVYKVIYGRFETPEEARAKEE 328  
 RESULT 11  
 ID Q55489 PRELIMINARY; PRT; 948 AA.  
 AC Q55489;  
 DT 01-NOV-1996 (T-EMBLrel. 01, Created)  
 DT 01-NOV-1996 (T-EMBLrel. 01, Last sequence update)  
 DT 01-MAY-1999 (T-EMBLrel. 10, Last annotation update)  
 DE HYPOTHETICAL 104.9 KD PROTEIN.  
 OS Synechocystis sp. (strain PCC 6803).  
 OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-PCC6803;  
 RA TABATA S.;  
 RL Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-PCC6803;  
 RX MEDLINE; 96127529.  
 RA KANEKO T., TANAKA A., SATO S., KOTANI H., SAZUKA T., MIYAJIMA N.,  
 RA SUGIURA M., TABATA S.;  
 RT "Sequence analysis of the genome of the unicellular cyanobacterium  
 Synechocystis sp. strain PCC6803. I. Sequence features in the 1 Mb  
 region from map positions 64% to 92% of the genome.";  
 RL DNA Res. 2:153-166(1995).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-PCC6803;  
 RX MEDLINE; 97061201.  
 RA KANEKO T., SATO S., KOTANI H., TANAKA A., ASAMIZU E., NAKAMURA Y.,  
 RA MIYAJIMA N., HIROSAWA M., SUGIURA M., SASAMOTO S., KIMURA T.,  
 RA HOSOUCHI T., MATSUNO A., MURAKI A., NAKAZAKI N., NARUO K., OKUMURA S.,  
 RA SHIMPO S., TAKEUCHI C., WADA T., WATANABE A., YAMADA M., YASUDA M.,  
 RA TABATA S.;  
 RT "Sequence analysis of the genome of the unicellular cyanobacterium  
 Synechocystis sp. strain PCC6803. II. Sequence determination of the  
 entire genome and assignment of potential protein-coding regions.";  
 RL DNA Res. 3:109-136(1996).  
 DR EMBL; D64006; BAA10838.1; -.  
 DR PFAM; PF00515; TPR; 5.  
 KW Hypothetical protein.  
 SQ SEQUENCE 948 AA; 104851 MW; 698807BB CRC32;

Query Match 6.0%; Score 189; DB 2; Length 948;  
 Best Local Similarity 22.1%; Pred. No. 0.00063;  
 Matches 118; Conservative 74; Mismatches 176; Indels 166; Gaps 22;  
 QY 2 SGLVLGRDEPAGHRL-----SQEILGSTRIVSQGLEALRSEHQVLSQTSQTECLQQG 57  
 Db 19 AGILINQID-PQGYGVVVTNKSQVTKYKL-----QIILPDANQLVEI---Q 62  
 QY 58 GHEGLVHEKARQLRRSMENI-----ELGLSEAOVMALASHLSTVES-- 100  
 Db 63 GH--QSLIKAQAMMDGSGSKESLLQAIELFTQGLEFYRQIGIVNRETVPAL-NRLGTIYSD 120  
 QY 101 -EKOKLRAQVRRLCQENQWLKDE--LAGTQORLQRSEQAQLEEKHLEFLQG----LR 154  
 Db 121 GEKSOALAYYQAITPLAQLODKALEGA--TLNNGIYINALADKKKAIDFYQOALVLR 178  
 QY 155 QYDEGHTSEEKGDAKDSLDLDFPNEEEEDPSNGLSRGGGATAAQGGYEPARLRTL 214  
 Db 179 Q-----AGDKTEE-----MTTI 190











GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 15, 2000, 00:39:17 ; Search time 2177.55 Seconds  
(without alignments)  
2010.452 Million cell updates/sec

Title: US-09-036-614A-2  
Perfect score: 2453  
Sequence: 1 GTGAAGTGGTGGAAGAGGG.....AGATCAGTCTTNTNTNNG 2453

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0  
Searched: 972840 seqs, 892348106 residues 1945680  
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0  
Maximum DB seq length: 1000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl:\*\*  
1: gb\_ba1:\*\*  
2: gb\_ba2:\*\*  
3: gb\_om:\*\*  
4: gb\_ov:\*\*  
5: gb\_pat:\*\*  
6: gb\_ph:\*\*  
7: gb\_pl1:\*\*  
8: gb\_pl2:\*\*  
9: gb\_pr1:\*\*  
10: gb\_pr2:\*\*  
11: gb\_pr3:\*\*  
12: gb\_ro:\*\*  
13: gb\_sts:\*\*  
14: gb\_sy:\*\*  
15: gb\_un:\*\*  
16: em\_fun:\*\*  
17: em\_hum1:\*\*  
18: em\_hum2:\*\*  
19: em\_in:\*\*  
20: em\_om:\*\*  
21: em\_or:\*\*  
22: em\_ov:\*\*  
23: em\_pat:\*\*  
24: em\_ph:\*\*  
25: em\_pl:\*\*  
26: em\_ro:\*\*  
27: em\_sts:\*\*  
28: em\_sy:\*\*  
29: em\_un:\*\*  
30: em\_v1:\*\*  
31: gb\_htg1:\*\*  
32: gb\_htg2:\*\*  
33: gb\_in1:\*\*  
34: gb\_in2:\*\*  
35: em\_ba1:\*\*  
36: em\_ba2:\*\*  
37: em\_hum3:\*\*  
38: em\_hum4:\*\*  
39: gb\_pr4:\*\*  
40: gb\_htg3:\*\*  
41: gb\_htg4:\*\*  
42: gb\_htg5:\*\*  
43: gb\_htg6:\*\*

44: gb\_htg7:\*\*  
45: em\_htg1:\*\*  
46: em\_htg2:\*\*  
47: em\_htg3:\*\*  
48: em\_hum5:\*\*  
49: gb\_pl3:\*\*  
50: gb\_pr5:\*\*  
51: gb\_htg8:\*\*  
52: gb\_htg9:\*\*  
53: gb\_htg10:\*\*  
54: gb\_htg11:\*\*  
55: gb\_htg12:\*\*  
56: gb\_htg13:\*\*  
57: gb\_htg14:\*\*  
58: gb\_in3:\*\*  
59: gb\_htg15:\*\*  
60: gb\_htg16:\*\*  
61: gb\_htg17:\*\*  
62: em\_htg4:\*\*  
63: em\_htg5:\*\*  
64: em\_htg6:\*\*  
65: em\_htg7:\*\*  
66: em\_hum6:\*\*  
67: gb\_htg18:\*\*  
68: gb\_htg19:\*\*  
69: gb\_htg20:\*\*  
70: gb\_htg21:\*\*  
71: gb\_htg22:\*\*  
72: gb\_htg23:\*\*  
73: gb\_htg24:\*\*  
74: gb\_htg25:\*\*  
75: gb\_htg26:\*\*  
76: gb\_htg27:\*\*  
77: gb\_htg28:\*\*  
78: gb\_htg29:\*\*  
79: gb\_htg30:\*\*  
80: gb\_htg31:\*\*  
81: gb\_vil:\*\*  
82: gb\_vil2:\*\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	1479.4	60.3	1524	11	HS0802612	AL162078 Homo sapi
2	841.2	34.3	2840	12	AF055666	AF055666 Mus muscu
3	785.4	32.0	1917	12	RATKINLB	M75147 Rat kinesin
4	785.4	32.0	2386	12	RATKINLC	M75148 Rat kinesin
5	783.8	32.0	2308	12	RATKINLA	M75146 Rat kinesin
6	762.8	31.1	2055	12	CGKINID	Y14586 Cricketulus
7	721.2	29.4	2289	12	AF055665	AF055665 Mus muscu
8	713.4	29.1	2308	9	HUMKINESLC	L04733 Homo sapien
9	593.4	24.2	2601	33	SUSKINLCIA	L10233 Strongyloce
10	593.4	24.2	2685	33	SUSKINLCIB	L10234 Strongyloce
11	593.4	24.2	2712	33	SUSKINLCIC	L10235 Strongyloce
12	571.8	23.3	1846	33	DROKINLIGH	L11013 Drosophila
13	571.8	22.9	9915	14	AF055298	L11328 Drosophila
14	560.6	22.9	9915	14	AF055298	AF055298 Expressio
15	540.8	22.0	2043	33	SQDSKLC	L24440 Loligo peal
16	532.8	21.7	2922	33	SUSKINLCID	L08258 Strongyloce
17	507.4	20.7	3089	4	GGU48359	U48359 Gallus gall
18	483	19.7	1930	33	SQDSKLCIA	L24441 Loligo peal
19	404.4	16.5	2035	33	CEKINLC1	229644 C.elegans m
20	403.8	16.5	1947	33	CEKINLC2	229645 C.elegans m
21	332	13.5	43154	34	CELC18C4	U55369 Caenorhabdi
22	299.6	12.2	95961	42	AC017329	AC017329 Drosophil
23	299.6	12.2	297385	34	AE003540	AE003540 Drosophil
24	291.2	11.9	90514	32	AL136304	AL136304 Homo sapi

```
25 224.2 9.1 23470 33 CEM7 268337 Caenorhabdi
26 224.2 9.1 84118 31 CEY3F2 298871 Caenorhabdi
27 160.4 6.5 1844 2 PBUT78597 78597 Plectonema
28 148.2 6.0 16440 51 AC023677 AC023677 Drosophila
29 146 6.0 64412 31 AP000630 AP000630 Homo sapi
30 146 6.0 164556 31 AP000759 AP000759 Homo sapi
31 146 6.0 185461 31 AP001107 AP001107 Homo sapi
32 145 5.9 186908 31 AP001812 AP001812 Homo sapi
33 142.4 5.8 4623 12 CRUKLCG1 L47236 Cricetus
34 140 5.7 32595 12 MUSXPDG1 L47235 Mus musculu
35 139.8 5.7 54336 11 HUMXPDG1 L47234 Homo sapien
36 139.8 5.7 160069 54 AC023329 AC023329 Homo sapi
37 139.2 5.7 205054 40 CNS01DXG AL39300 Homo sapi
38 120 4.9 85780 56 AC023803 AC023803 Mus muscu
39 77.8 3.2 660 8 HNS01DX3 AL116341 Botrytis
40 77.8 3.2 1150 81 HS4ULIR3 J02079 epstein-bar
41 77.8 3.2 5452 14 U02454 U02454 Cloning vec
42 77.8 3.2 9600 5 A92665 A92665 Sequence 1
43 77.8 3.2 10596 5 I25041 I25041 Sequence 15
44 77.8 3.2 10596 5 I30503 I30503 Sequence 15
45 77.8 3.2 10737 14 XU02428 XU02428 Cloning vec
```

## ALIGNMENTS

```
RESULT 1
HSM802612 1524 bp mRNA 23-MAR-2000
LOCUS Homo sapiens mRNA; CDNA DKFzp761H229 (from clone DKFzp761H229);
DEFINITION partial cds.
ACCESSION AL162078
VERSION AL162078.1 GI:7328159
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1524)
AUTHORS Blum.H., Bauersachs.S., Mewes.H.W., Weil.B. and Wiemann.S.
TITLE Direct Submission
JOURNAL Submitted (15-MAR-2000) MIPS, Am Klopferspitz 18a, D-82152
Martinsried, GERMANY
COMMENT Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center-(DKFZ); Email s.wiemann@dkfz-heidelberg.de;
sequenced by LMU (Ludwig Maximilians University,
Munich/Germany) within the cDNA sequencing consortium of the German
Genome Project.
This clone (DKFzp761H229) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further
information about the clone and the sequencing project is available
at http://www.mips.biochem.mpg.de/proj/cdna/.
Location/Qualifiers
1..1524
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="DKFzp761H229"
/clone_lib="761 (synonym: hamy2). Vector psportil; host
DH10B; sites Noti + SalI"
/dev_stage="adult"
/tissue_type="amygdala"
1..1068
/gene="DKFzp761H229"
<1..1068
/gene="DKFzp761H229"
/note="similarity to kinesin light chain (Homo sapiens)"
/codon_start=1
/product="hypothetical protein"
/protein_id="CA882411.1"
/db_xref="GI:7328160"
/translation="DONKYKEAAHLNDALSIRETLGPDHPAVAATLNNLAVLYGKR
GRYKEAPLCORALETRKVLGTNHPDVAKQLNNLALLCQNOGKVEAYERYQALAI
```

## FEATURES

```
source
1..1524
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="DKFzp761H229"
/clone_lib="761 (synonym: hamy2). Vector psportil; host
DH10B; sites Noti + SalI"
/dev_stage="adult"
/tissue_type="amygdala"
1..1068
/gene="DKFzp761H229"
<1..1068
/gene="DKFzp761H229"
/note="similarity to kinesin light chain (Homo sapiens)"
/codon_start=1
/product="hypothetical protein"
/protein_id="CA882411.1"
/db_xref="GI:7328160"
/translation="DONKYKEAAHLNDALSIRETLGPDHPAVAATLNNLAVLYGKR
GRYKEAPLCORALETRKVLGTNHPDVAKQLNNLALLCQNOGKVEAYERYQALAI
```

```
YEQGLPDNPVARTKNNLASCYLKQGYAEATLYKEILTRAHVQFSGVDDDKPI
WHAEEEREMSKSRHHEGTPYAEYGMWYKACKVSSPTVNTLNLGALYRQCKLEA
AETLEECALRSRROGTPISQTKVAELLGSDGRRTSQEGPDSVFEEGEDASVAVE
WSGDSGTLOBSGSLGKIRDLVLRSSSELLVRKLOGTEPRFSSSMKRAASLINTINOPS
AAPQVSRGUSASTMDLSSSS"
polyA_signal 1468..1473
polyA_site 1489
BASE COUNT 382 a 421 c 436 g 285 t
ORIGIN
Query Match 60.3%; Score 1479.4; DB 11; Length 1524;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1480; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 961 GACCAAGATAAGTAAGCAAGCTGCCACCTGCTGTAATGATGCCCTTAGCATCCGGAG 1020
Db 1 GACCAAGATAAGTAAGCAAGCTGCCACCTGCTGTAATGATGCCCTTAGCATCCGGAG 60
QY 1021 AGCACCTTGGGACCTGACCATCTCTGCTGGCTGCCACCTCAACAATTTGGCTGTGCTC 1080
Db 61 AGCACCTTGGGACCTGACCATCTCTGCTGGCTGCCACCTCAACAATTTGGCTGTGCTC 120
QY 1081 TATGCAAAAGGGCAAGTACAAGGAGGACAGAGCTCTGTGCCAGCGGCACTGGAGATT 1140
Db 121 TATGCAAAAGGGCAAGTACAAGGAGGACAGAGCTCTGTGCCAGCGGCACTGGAGATT 180
QY 1141 CGAGAAAAGGTCCTGGGCAACGTAATCCAGATGTGCAAAACAGCTGAACAACCTGGCC 1200
Db 181 CGAGAAAAGGTCCTGGGCAACGTAATCCAGATGTGCAAAACAGCTGAACAACCTGGCC 240
QY 1201 CTCCTTGTGCCAAAACGAGGCAAGTATGAGGCGCTGGAACGCTACTACCAAGGAGACTG 1260
Db 241 CTCCTTGTGCCAAAACGAGGCAAGTATGAGGCGCTGGAACGCTACTACCAAGGAGACTG 300
QY 1261 GCCATCTACGAGGGGACCTGGGCGGCAACACCTTAATGTAGCCCGGCAACAGAACAC 1320
Db 301 GCCATCTACGAGGGGACCTGGGCGGCAACACCTTAATGTAGCCCGGCAACAGAACAC 360
QY 1321 CTGGCTTCTCTTACCTGAAACAGGCAAAATATCTGAGGCTGAGACACTATACAAGAG 1380
Db 361 CTGGCTTCTCTTACCTGAAACAGGCAAAATATCTGAGGCTGAGACACTATACAAGAG 420
QY 1381 ATCTGACCCCTGCCATGTACAGGAGTTTGGGCTCTGTGGATGATGACCAAGCCCATC 1440
Db 421 ATCTGACCCCTGCCATGTACAGGAGTTTGGGCTCTGTGGATGATGACCAAGCCCATC 480
QY 1441 TGGATGATGACAGAGGAGGAGGAGAAATGAGCAAAAGCCGACCATGAGGGTGGGACA 1500
Db 481 TGGATGATGACAGAGGAGGAGGAGAAATGAGCAAAAGCCGACCATGAGGGTGGGACA 540
QY 1501 CCCTATGCTGAGTATGAGGCTGTACAAGGCTGCAAGGCTGCAAGGCTGCAAGGCTGCA 1560
Db 541 CCCTATGCTGAGTATGAGGCTGTACAAGGCTGCAAGGCTGCAAGGCTGCAAGGCTGCA 600
QY 1561 ACTACTCTGAGAAACCTGGGAGCTCTGTATAGGCGCCAGGAGAAAGCTGGAGGTGCTGAG 1620
Db 601 ACTACTCTGAGAAACCTGGGAGCTCTGTATAGGCGCCAGGAGAAAGCTGGAGGTGCTGAG 660
QY 1621 ACCCTGGAGGAATGTGCCCTCGGTCGGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1680
Db 661 ACCCTGGAGGAATGTGCCCTCGGTCGGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGG 720
QY 1681 AAGGTGGCAGAGCTGCTTGGGAGAGTGTAGTGAAGGAGCTCCCAAGGAGGAGGAGGAGG 1740
Db 721 AAGGTGGCAGAGCTGCTTGGGAGAGTGTAGTGAAGGAGCTCCCAAGGAGGAGGAGGAGG 780
QY 1741 GACAGTGTGAAATTCGAGGGTGTGGAAGATGCTTCTGTGCTGTGGAGTGTGCTGGGGAT 1800
Db 781 GACAGTGTGAAATTCGAGGGAGGTGGAAGATGCTTCTGTGCTGTGGAGTGTGCTGGGGAT 840
QY 1801 GGCAGTGGGACCTCGACAGAGGAGTGGCTCTCTTGGCAAGATCGGGGATGTGCTCCGAGA 1860
Db 1801 GGCAGTGGGACCTCGACAGAGGAGTGGCTCTCTTGGCAAGATCGGGGATGTGCTCCGAGA 1860
```



Db 1857 TCCTTTAACTCTCTTAAACAAGAGTGTGGAGAGCCAGTCCAG 1898

RESULT 3

RATKINLB RATKINLB 1917 bp mRNA ROD 27-APR-1993

LOCUS Rat kinesin light chain B mRNA.

DEFINITION M75147

ACCESSION M75147.1 GI:205079

VERSION 1

KEYWORDS kinesin light chain.

SOURCE Rattus norvegicus Adult Brain cDNA to mRNA.

ORGANISM Rattus norvegicus

REFERENCE 1 (bases 1 to 1917)

AUTHORS Cyr, J.L., Pfister, K.K., Bloom, G.S., Slaughter, C.A. and Brady, S.T.

TITLE Molecular genetics of kinesin light chains: Generation of isoforms by alternative splicing

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 88, 10114-10118 (1991)

MEDLINE 92052221

FEATURES

Location/Qualifiers

1..1917

source

1..1917

organism="Rattus norvegicus"

/db\_xref="taxon:10116"

/dev\_stage="Adult"

/tissue\_type="Brain"

BASE COUNT 491 a 496 c 608 g 322 t

ORIGIN

Query Match 32.0%; Score 785.4; DB 12; Length 1917;

Best Local Similarity 71.7%; Pred. No. 3.2e-176;

Matches 1074; Conservative 0; Mismatches 411; Indels 12; Gaps 3;

QY 216 GCTCAGCAGAGGAGATCTGGGGAGCACACGGCTGGTCCAGCCAGGGCTAGAGGCCCT 275

Db 188 GCTTACGAGGATGAATCATTTCTAAGACCAGCAGGTGATCCAGGGCTGGAGGCCCT 247

QY 276 AGCAGTGAACACACCGGCGCTGTGCAAGCCTGTCCAGCACCATTGAGTCTCTCAGCA 335

Db 248 GAAGAATGAGCAACCTCCATCTCCAGAGTCTGTGGAGCGCTGAAGTCTGTGAAGA 307

QY 336 GGAGGCCATGAGGAGGCTGTGTCATGAGAAGCCGCGAGCTTCGCGTCTCATGGA 395

Db 308 G---GAGCATGAGAGCAACCTGTGGAGGAGAGTCCAGCATGATCCGCAAGTCTTGA 364

QY 396 AAACATTTAGCTCGGGCTGAGTGAGGCCAGGTGATGCTGTAGCCAGCCACTGAG 455

Db 365 GATGCTGGAGCTGGCTGAGGAGCGCAGGTGATGATGCGCCCTGTCTAATCACTGAA 424

QY 456 CACAGTGGAGTGGGAGAAACAGAGCTGGGGCTAGGTGGGGGGCTATGCCAGAGAA 515

Db 425 TGCGGTGGAGTCCGAGAGCAAGCTCCGAGCTCAGGTTTCAGCTGTGTGCCAGGAA 484

QY 516 CCAGTGGCTGGGAGTGAAGTGGCTGGCCAGCAGCGGCTACAGCGCAGTGAACAGCG 575

Db 485 CCAGTGGCTGGGAGTGAAGTGGCTGGCCAGCAGCGGCTACAGCGCAGTGAACAGCG 544

QY 576 TGTGGCTCAGCTGGAGAGGAAAGACCTGGAGTTCCTGGGGCAGCTGCGGCAGTA 635

Db 545 CGTGGCTCAGCTGGAGAGGAGAAAGCAGCTGGAGTTCATGAACACAGCTGAAGAATA 604

QY 636 TGATGAGGATGGACATACCTCCGAGGAGAAAGCGGATGCCCAAGATTCCTCGA 695

Db 605 CGAGCAGGACATCTCTCCCTCGGAGAGCAAGAGACTCTGATTTCTTCCAAAGAGCCGTGA 664

QY 696 TGACCTCTTTCTTAATGAGGAGAAAG---GGACCCAGCAATGGCTTGTCCCGTGGTCA 752

Db 565 TGACCTCTTCCGAGTGAAGTGAAGTGAACCCAGGAGCAAGGATCCAGCAACACAGCAG 724

QY 753 AGGTGCTACAGAGCTCAGCAGGGTGGATATGATATCCAGCAAGGTTTCGGACAGTTGA 812

Db 725 TGCTGCGCGCGCGCCAGCAGCGGCGGCTACGAGATCCCTGCCCGGCTGCGCAGCTCCA 784

Db 1857 TCCTTTAACTCTCTTAAACAAGAGTGTGGAGAGCCAGTCCAG 1898

## RESULT 3

RATKINLB RATKINLB 1917 bp mRNA ROD 27-APR-1993

LOCUS Rat kinesin light chain B mRNA.

DEFINITION M75147

ACCESSION M75147.1 GI:205079

VERSION 1

KEYWORDS kinesin light chain.

SOURCE Rattus norvegicus Adult Brain cDNA to mRNA.

ORGANISM Rattus norvegicus

REFERENCE 1 (bases 1 to 1917)

AUTHORS Cyr, J.L., Pfister, K.K., Bloom, G.S., Slaughter, C.A. and Brady, S.T.

TITLE Molecular genetics of kinesin light chains: Generation of isoforms

by alternative splicing

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 88, 10114-10118 (1991)

MEDLINE 92052221

FEATURES

Location/Qualifiers

1..1917

source

1..1917

organism="Rattus norvegicus"

/db\_xref="taxon:10116"

/dev\_stage="Adult"

/tissue\_type="Brain"

BASE COUNT 491 a 496 c 608 g 322 t

ORIGIN

Query Match 32.0%; Score 785.4; DB 12; Length 1917;

Best Local Similarity 71.7%; Pred. No. 3.2e-176;

Matches 1074; Conservative 0; Mismatches 411; Indels 12; Gaps 3;

QY 216 GCTCAGCAGAGGAGATCTGGGGAGCACACGGCTGGTCCAGCCAGGGCTAGAGGCCCT 275

Db 188 GCTTACGAGGATGAATCATTTCTAAGACCAGCAGGTGATCCAGGGCTGGAGGCCCT 247

QY 276 AGCAGTGAACACACCGGCGCTGTGCAAGCCTGTCCAGCACCATTGAGTCTCTCAGCA 335

Db 248 GAAGAATGAGCAACCTCCATCTCCAGAGTCTGTGGAGCGCTGAAGTCTGTGAAGA 307

QY 336 GGAGGCCATGAGGAGGCTGTGTCATGAGAAGCCGCGAGCTTCGCGTCTCATGGA 395

Db 308 G---GAGCATGAGAGCAACCTGTGGAGGAGAGTCCAGCATGATCCGCAAGTCTTGA 364

QY 396 AAACATTTAGCTCGGGCTGAGTGAGGCCAGGTGATGCTGTAGCCAGCCACTGAG 455

Db 365 GATGCTGGAGCTGGCTGAGGAGCGCAGGTGATGATGCGCCCTGTCTAATCACTGAA 424

QY 456 CACAGTGGAGTGGGAGAAACAGAGCTGGGGCTAGGTGGGGGGCTATGCCAGAGAA 515

Db 425 TGCGGTGGAGTCCGAGAGCAAGCTCCGAGCTCAGGTTTCAGCTGTGTGCCAGGAA 484

QY 516 CCAGTGGCTGGGAGTGAAGTGGCTGGCCAGCAGCGGCTACAGCGCAGTGAACAGCG 575

Db 485 CCAGTGGCTGGGAGTGAAGTGGCTGGCCAGCAGCGGCTACAGCGCAGTGAACAGCG 544

QY 576 TGTGGCTCAGCTGGAGAGGAAAGACCTGGAGTTCCTGGGGCAGCTGCGGCAGTA 635

Db 545 CGTGGCTCAGCTGGAGAGGAGAAAGCAGCTGGAGTTCATGAACACAGCTGAAGAATA 604

QY 636 TGATGAGGATGGACATACCTCCGAGGAGAAAGCGGATGCCCAAGATTCCTCGA 695

Db 605 CGAGCAGGACATCTCTCCCTCGGAGAGCAAGAGACTCTGATTTCTTCCAAAGAGCCGTGA 664

QY 696 TGACCTCTTTCTTAATGAGGAGAAAG---GGACCCAGCAATGGCTTGTCCCGTGGTCA 752

Db 565 TGACCTCTTCCGAGTGAAGTGAAGTGAACCCAGGAGCAAGGATCCAGCAACACAGCAG 724

QY 753 AGGTGCTACAGAGCTCAGCAGGGTGGATATGATATCCAGCAAGGTTTCGGACAGTTGA 812

Db 725 TGCTGCGCGCGCGCCAGCAGCGGCGGCTACGAGATCCCTGCCCGGCTGCGCAGCTCCA 784

QY 813 CAACCTGGTATCCAGTACGAGCCCAAGGTGCTGATGAGTGGCCCTGCCATCTGTGTA 872  
Db 785 CAACCTGGTATCCAGTATGCTCGCAGGGCGTTACGAGTGGCGTGCACCTGTGCAA 844  
QY 873 GCAGGCACTAGAGCACTGGAGCCACATCAGCGCGTGGCCACCTGATGTCGCCACCAT 932  
Db 845 GCAGGCGCTGGAGCACTGGAGAGACTTCGGCCAGCACCTGATGTCGCCACCAT 904  
QY 933 GCTCAACATCCTGCTTTGGTGTATCGTACCAAGATTAAGTAAGAACTGCCACCT 992  
Db 905 GCTCAACATCCTGCTGCTGCTGATGAGGATCAGAACAGTATAAAGACSCACCAACCT 964  
QY 993 GCTGAATGATGCTTATAGCTCCGGGAGAGCACTTGGGACCTGACCATCCTGCTGTGGC 1052  
Db 965 CCTGAACGATGCTGCTGCTATCCGTGAGAAACCTCGCGCGAGATCACCTCGGTGGC 1024  
QY 1053 TGCCACACTCAACAATTTGGTGTGCTATGCAAAAGGGCAAGTACAAAGAGAGCAGA 1112  
Db 1025 TGCACCTCAACAACCTAGCAGTGTGATGAGTACGGTGGCAAGTACAAAGAGGCCGA 1084  
QY 1113 GCCTCTGTGCGAGCGGCACTGGAGATTCGAGAAAGGTCTCGGCACGCAATCATCCAGA 1172  
Db 1085 GCCCTGTGCAACGGCCCTGGAGATCAGGAGAGGTTCTGGAAAGGATCACCTCTGA 1144  
QY 1173 TGTGGCAAAACAGCTGACAACTGGCCCTCTGTCGCAAAACAGGGCAAGTATGAGGC 1232  
Db 1145 TGTGCTAAGCACTAATATCTGGCTTGTCTGTCAGAACCCAGGGCAAGTACGAGGA 1204  
QY 1233 CGTGAACGCTACTACCAAGGAGCACTGGCATCTACGAGGGGCAAGTGGGCGCGACAA 1292  
Db 1205 GGTGAGATTAATACCAAGAGGGCCCTGGAGATCTACCAAGCAAGTCTGGACCCGATGA 1264  
QY 1293 CCCTAATGTAGCCCGGACCAACCACTGGCTTCTGTTTACCTGAAACAGGGCAAAATA 1352  
Db 1265 TCCCAACGTGSCCAAGACGAAGATAACCTGGCTTCTGTTTATCTGAAACAAGGAGTT 1324  
QY 1353 TGCTGAGGCTGAGACATATACAAGAGATCTGACCCGCTGCCATGTACAGAGTTGG 1412  
Db 1325 CAACAAGCAGAAACATTTGACAAAGAGATCTCACCCGCGCATGAAACGGAGTTGG 1384  
QY 1413 GTCTGTGGATGATCACCACAGCCCATCTGATGTCATGACAGAGAGCGGGAGAAATGAG 1472  
Db 1385 ATCTGTGGATGATGAACAAGCCCATATGATGTCACCGGAGAGAGAGAGGTGCAA 1444  
QY 1473 CAAGAAGCGGCAACATGAGGGTGGACACCTATGCTGATGAGGCTGGTACAAAGGC 1532  
Db 1445 AGGAA-----AGCAGAAGGAGCGGTCTGTTTGGAGAGTATGGCGCTGTTAAGGC 1498  
QY 1533 CTGCAAGTGAAGAGCGCCACAGTGAACACTACTCTGAGAAACCTGGAGCTCTGTATAG 1592  
Db 1499 CTGCAAGTGGACAGTCCCAACCGCTCAACAACACCTTGAACCAAGCTGGGACACTTTACCG 1558  
QY 1593 GCGCCAGGGAAGCTGGAGGCTGTGAGACCTGGAGGAATGTCCTGCGGTCCCGGAG 1652  
Db 1559 CCGCCAGGGAAGTTGAGCTGGGAGACTGGAGGAAGCTGCCCTGAGGTACGCTAA 1618  
QY 1653 ACAGGCACTGACCTATACCCAGAGAAAGTGGCAGAGCTGTTGGGAGAGTGA 1709  
Db 1619 GCAGGCTTTGACAATGTTCAACAACAGAGAGTGGCTGAAGTGTCTAAATGACCCCTGA 1675

## RESULT 4

RATKINLC  
LOCUS RATKINLC 2386 bp mRNA  
DEFINITION Rat kinesin light chain C mRNA.  
ACCESSION M75148  
VERSION M75148.1 GI:205080  
KEYWORDS kinesin light chain.  
SOURCE Rattus norvegicus Adult Brain cdna to mRNA.  
ORGANISM Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
FEATURES  
source

1 (bases 1 to 2386)  
Cyr,J.L., Pfister,K., Bloom,G.S., Slaughter,C.A. and Brady,S.T.  
Molecular genetics of kinesin light chains: Generation of isoforms  
by alternative splicing  
Proc. Natl. Acad. Sci. U.S.A. 88, 10114-10118 (1991)  
92052221 Location/Qualifiers  
1..2386  
/organism="Rattus norvegicus"  
/db\_xref="taxon:10116"  
/dev\_stage="Adult"  
/tissue\_type="Brain"

BASE COUNT 612 a 588 c 722 g 464 t  
ORIGIN

Query Match 32.0%; Score 785.4; DB 12; Length 2386;  
Best Local Similarity 71.7%; Pred. No. 3.1e-176;  
Matches 1074; Conservative 0; Mismatches 411; Indels 12; Gaps 3;

QY 216 GCTACGCAAGAGAGAGATCTCTGGGAGCACACAGCGCTGCTCAGCCAAAGGCTAGAGGCCCT 275  
Db 174 GCTTACGCGAGATGAATCAATTTCTAAGACCAAGCAGGTGATCCAGGGCTTGGAGGCCCT 233  
QY 276 AGCGAGTGAACACAGCGCGCTGCTGCAAAAGCTGTCCAGACCATTTGAGTGTCTGCAGCA 335  
Db 234 GAAGAATGAGCACAACTCCATCTCGAGAGTCTGCTGGAGACGCTGAAGTCTTGAAGAA 293  
QY 336 GGGAGCCATGAGGAGGCTGGTGCATGAGAGGCCCGGCGAGCTTCGCCCTTCTATGGA 395  
Db 294 G---GACATGAGAGCAACCTGGTGGAGGAAGTCCAGCATGATCCGCAAGTCTCTTGA 350  
QY 396 AAACATTTAGGCTCGGCGCTGAGTGGGCCCGAGGTGATGTGCTGTAGCACCCACCTTGAG 455  
Db 351 GATGCTGGAGCTCGGCTTGGAGGCGCAGGTGATGATGSCCTGTCTATACCTTGAA 410  
QY 456 CACAGTGGAGTCGGAGAACAGAGCTCGGGCTCAGTGGCGGCTATCCAGGAGAA 515  
Db 411 TGCCCTGGAGTCCGAGAGCAGAGCTCCGAGCTCAGTTCGAGCTGTGCCAGGAA 470  
QY 516 CCAGTGGCTGGGAGTGGCTGGTGGCCACCCAGCGGCTACAGCCAGTGCAGCAGGC 575  
Db 471 CCAGTGGCTGGGAGTGAACCTGGCCACACAGCAGAAAGTTTACAGAGAGCGAGCATC 530  
QY 576 TGTGCTCAGCTGGAGGAGAAAGCACTGTGAGTCTCTGGGAGCTGCGGCGAGTA 635  
Db 531 CGTGCTCAGCTGGAGGAGAGAGCAGCAGTGGAGTTTCAATGACCAAGCTTGAAGATA 590  
QY 636 TGATGAGTGGACATACCTCGAGGAGAAAGAGCGGATGCCACCAAGGATTCCTCTGGA 695  
Db 591 CGACGACGACATCTCCCTCGGAAGACAAAGACTCTGATTTCCAAAGAGCGGTTGA 650  
QY 696 TGACCTCTTCTTATGAGGAGGAGAA---GGACCCCAAGCAATGGCTGTCCCTGGTCA 752  
Db 651 TGACCTCTTCCGGAATGATGAGGATGACCCAGGACAAAGGAATCCAGCAGCAACAGCAG 710  
QY 753 AGGTGCTACAGCAGCTCAGCAGGTTGGATATGAGATCCAGCAAGGTTGCGGAGCTTGA 812  
Db 711 TGCTGCGGCCGCCCGCAGAGGGCGGTACGAGATCCCTGCCCGGTGCGCAGCTTCCA 770  
QY 813 CAACCTGTGTATCCAGTACGCGCCCAAGGTGCGTATGAGTGGCGGCTGCCACTCTGTAA 872  
Db 771 CAACCTGTGTATCCAGTATGCTCGCAGGGCGGTTACGAGGTGGCGGTGCCACTCTGCA 830  
QY 873 GCAGGCACTAGAGGACCTGGAGCGCACATCAGGCGGTGGCCACCTGATGTGCGCCACAT 932  
Db 831 GCAGGCGCTGGAGGACCTGGAGAGAGACTTCCGGCCACGACCCCTGATGTGGCCACCAT 890  
QY 933 GCTCAACATCCTGCTTTGGTGTATCGTGACAGAAATAGTATAGGAAGCTGCCACCT 992  
Db 891 GCTCAACATCCTGGCCCTGGTGTACAGGGATCAGAAACAGTATATAAGACGAGCAACCT 950  
QY 993 GCTGAATGATGCCCTTAGCATCCGGGAGAGCAGCTTGGGACCTGACCATCCTGCTGTGGC 1052

```
Db 951 CCTGAACGATCCCTGGCTATCCCTGGATGAGAAACCCCTGGCGAGATCACCTCGGTGGC 1010
Qy 1053 TGCACACTCAACAAATTTGGCTGTCTATGGCAAAAGGGGCAAGTACAAAGAGGCGAGA 1112
Db 1011 TGCACCCCTCAACAACTAGCAGTGTCTAGCGTAAAGCGGTGGCAAGTACAAAGAGCGCGA 1070
Qy 1113 GCCTCTGTGCAGCGGGCACTGGAGATTCGAGAAAGGTCCTGGCGACGAAATCATCCAGA 1172
Db 1071 GCCGCTGTGCAAAACGGGCGCTGGAGATCAGGGAGAAGTTCTGGGAAAGGATCACCCCTGA 1130
Qy 1173 TGTGGCAAAACAGCTGACACACCTGGCGCTCTGTGCAAAACAGGGCAAGTATGAGGC 1232
Db 1131 TGTGCTAAAGCAATTAATAATCTGGCTGTCTGTGCAAAACAGGGCAAGTACAGGA 1190
Qy 1233 CGTGAACGCTACTACACAGCGACACTGGCCATCTACGAGGGGCGAGCTGGGGCGGACAA 1292
Db 1191 GGTGGAGTATTACTACAGAGGGCCCTGGAGATCTACACAGNAAGCTCGGACCCGATGA 1250
Qy 1293 CCTAATGTAGCCGGACCAAGAACAACTGGCTTCTGTACTGTAAACAGGGCAAAATA 1352
Db 1251 TCCCAACGTGGCAAGACGAAAGATAACCTGGCTTCTGTATCTGAACAAGGAAGTT 1310
Qy 1353 TGCTGAGGCTGAGACATATACAAAGAGATCCTGACCCGTGCCATGTACAGAGTTGG 1412
Db 1311 CAACACAGCAGAAACATTTGTAAGAGGATTTCTCACCCGCGCATGAACGGGAGTTGG 1370
Qy 1413 GTCTGTGATGATGACCAACAGCCCTCTGGATCGATCGAGAGAGGGGAGGAAATGAG 1472
Db 1371 ATCTGTGATGATGAACACAGCCCATATGATGATCGACGCGGAGAGAGAGGTGCAA 1430
Qy 1473 CAAGAAGCGGCACCATGAGGTTGGGACACCCATGCTGTAGTATGAGGAGTGGTACAGGC 1532
Db 1431 AGGAA-----AGCAGAAGGACGGTGCTCTTTTGGAGAGTATGGCGCTGGTATAAGGC 1484
Qy 1533 CTGCAAGTGAAGCCCAAGTGAACACTCTCTGATCGATCGAGAGAGGGGAGGAAATGAG 1592
Db 1485 CTGCAAGTGAAGTTCACCCCGTCAACACCCCTTGAACACCTGGGAGCATTTACCG 1544
Qy 1593 CGCCAGGGAAGCTGGAGGCTGTGAGACCTCGAGAGGAATGTGCCCTGGGTCGCGGAG 1652
Db 1545 CGCCAGGGAAGTTTGAAGCTGCGGAGACACTGGAGAGCTGSCCTGAGGTACAGTAA 1604
Qy 1653 ACAGGCACTGACCTATCAGCAGAGAGAGTGGCAGAGCTGTTGGGAGAGTGA 1709
Db 1605 CGAGGCTTGACAATGTTCAACAACAGAGAGTGGCTGAAGTGTCTAAATGACCCCTGA 1661
```

## RESULT 5

```
RATKINLA RATKINLA 2308 bp mRNA ROD 27-APR-1993
LOCUS Rat kinesin light chain A mRNA.
DEFINITION M75146
ACCESSION M75146
VERSION M75146.1 GI:205078
KEYWORDS kinesin light chain.
SOURCE Rattus norvegicus Adult Brain cDNA to mRNA.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
1 (bases 1 to 2308)
REFERENCE Cyt.J.L., Pfister,K.K., Bloom,G.S., Slaughter,C.A. and Brady,S.T.
AUTHORS Molecular genetics of kinesin light chains: Generation of isoforms
TITLE by alternative splicing
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 88, 10114-10118 (1991)
MEDLINE 92052221
FEATURES Location/Qualifiers
Source 1..2308
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
/dev_stage="Adult"
/tissue_type="Brain"
BASE COUNT 600 a 577 c 686 g 445 t
ORIGIN
```

Query Match 32.0%; Score 783.8; DB 12; Length 2308;  
Best Local Similarity 71.7%; Pred. No. 7.6e-176;  
Matches 1073; Conservative 0; Mismatches 412; Indels 12; Gaps 3;

```
Qy 216 GCTCAGCAAGAGAGATCTCTGGGAGACACAGGCTGGTCAAGCAAGGGCTAGAGGCCCT 275
Db 165 GCTTAGCAGGATGAATCTTCTTAAGCAACAGCAGGTGATCCAGGGGCTGGAGGCCCT 224
Qy 276 AGCAGTGAACACAGGCGCTGCTGCAAGGCTGCTCCAGACCAATGAGTGTCTGCAACA 335
Db 225 GAAGAATGAGCACAACACTCCATCTGACAGTCTGCTGGAGACGCTGAAGTGTCTTGAAGA 284
Qy 336 GGGAGGCATCAGGAAGGGCTGGTGCATGAGAAGGCCGCGCAGCTTCGCCCTTCTTATGGA 395
Db 285 G--CACGATGAGAGCAACCTGGTGGAGGAGAAGTCAGCATGATCCGCAAGTCTTTGGA 341
Qy 396 AAACATGAGTCTGGGCTGAGTGAAGGCCAGGTGATGCTGGCTCTAGCCAGCCACCTGAG 455
Db 342 GATGCTGAGCTCGGCCCTGAGCGAGGCGAGGTGATGATGCCCTGTCTAATCACCCTGAA 401
Qy 456 CACAGTGAAGTCGGAGAAACACAGAGCTCGCGGCTCAGTGCAGCGGCTATGCCAGGAGA 515
Db 402 TCCCTGGAGTCCGGAAGACAGAGTCCGAGCTCAGGTTGACGCTCTGTGCCAGGAGA 461
Qy 516 CCAGTGGCTGGGGATGAGCTGGCTGGCACCCAGCAGCGGTACAGGCGCATGTGAACAGGC 575
Db 462 CCAGTGGCTGGGGATGAACCTGGCCAACACACACAGAGAAGTTACAGAAGACAGCAGTC 521
Qy 576 TGTGCTCAGCTGGAGAGGAAAGAACACCTCGAGTTCCTGGGGCAGCTGCGGCGAGTA 635
Db 522 CGTGTGCTGAGTGGAGAGGAAAGAACACCTCGAGTTCATGAACCAAGCTGAAGAATA 581
Qy 636 TCATCAGATGACATACCTCGAGGAGAAAGAGGCGATGCCACCAAGGATTCCTCTGGA 695
Db 582 CGACAGCAGATCTCTCCCTCGGAAGACAAAGACTGTGATTTCTTCCAAAGAGCGCTTGA 641
Qy 696 TGACCTCTTTCTTAATAGAGGAGA--GGACCCCAAGCAATGGCTTGTCCCTGGTCA 752
Db 642 TGACCTCTTCCGATGATGAGGATGATGCCAGGACAAAGGAATCCAGCAGCAACACAG 701
Qy 753 AGGTCTACAGCAGCTCAGCAGGGTGGATATGATGCCACCAAGGTTGCGGAGCTTGCA 812
Db 702 TGCTCGCGCGCGCCAGCAGGCGGCTACGAGATCCCTGCCGCGCTGCGCAGCTCCA 761
Qy 813 CAACCTGTGTATCCAGTACGACGAGCCCAAGTCTGATGAGTGGCGGTGCCACTCTGTAA 872
Db 762 CAACCTGTGTATCCAGTATGCTTCGAGGGCGGTTACGAGTGGCGGTGCCACTCTGCAA 821
Qy 873 GCAGCACTAGAGGACCTGGAGCGCACATCAGGCGCTGGCCACCCCTGATGTGCGCACCAT 932
Db 822 GCAGGCGCTGGAGGACCTGGAGAAGACTTCGCGCCACGACACCACCTGATGTGGCCACCAT 881
Qy 933 GCTCAACATCTTGTGTTGGTGTATCTGTGACCAAGTAAGTATAGGAGTGGCCACCT 992
Db 882 GCTCAACATCTTGGCCCTGGTGTACAGGATCAGAACAAAGTATAAAGACGCGCAAACT 941
Qy 993 GCTGAATGATGCCCTTAGCATCCGAGGAGCAGCTTGGGACCTGACCATCTGCTGTGGC 1052
Db 942 CCGTGAACGATGCCCTGGGTATCCGTGAGAAAACCCCTGGCGCGAGATCACCTCGGGTGGC 1001
Qy 1053 TCCACACTCAACAAATTTGGCTGTCTTATGGCAAAAGGGGCAAGTACAAAGAGGCGAGA 1112
Db 1002 TCGCAGCCCTCAACAACTAGCAGTGTCTGACGTTAAGCGTGGCAAGTACAAAGAGGCGGA 1061
Qy 1113 GCCTGTGCGCAGCGGCGACTGGAGATTCGAGAAAAGTCTCTGGCAGGAATCATCAGA 1172
Db 1062 GCGGCTGTCAAAACGGGCCCTGGAGATCAGGGAAGGTTCTGGGAAAGGATACCCCTGA 1121
Qy 1173 TGTGGCAAAACAGCTGAACACCTGGCCCTCTTGTGCCAAAACCAAGGCAAGTATGAGGC 1232
Db 1122 TGTGCTAAGCAACTAATATATCTGGCCCTTGTGTGCCAAGAACCGGCAAGTACGAGGA 1181
```



QY 1233 CGTGAACGCTACTACAGGAGGAGCTGGCCATCTACAGAGGGGAGCTGGGGCCGGACAA 1292  
Db 1182 GGTGGAGTATTACTACAGAGGGCCCTGGAGATCTACCAAGAGCTCGGACCCGATGA 1241  
QY 1293 CCTAATGTAGCCGAGCAAGAACAACTGGCTTCCTGTTACCTGAAACAGGGCAATA 1352  
Db 1242 TCCCAACGTGGCAAGAGAGAAATACCTGGCTTCCTGTTATCTGAACAAGGAAGTT 1301  
QY 1353 TGCTGAGGCTGAGACACTATAAAGAGATCTGACCCGTCGCCATGTACAGGATTTGG 1412  
Db 1302 CAACAAGCAGAAACATTGTACAAGAGATTCTACCCGCGCACATGAACGGATTTGG 1361  
QY 1413 GTCTGTGATGATGACCAAGCCATCTGTGATGATGACAGAGCGGGAGGAATGAG 1472  
Db 1362 ATCTGTGATGATGAAACCAAGCCATATGATGACAGCGGAGGAGAGAGAGTCAA 1421  
QY 1473 CAAGAAGCCGACCATGAGGTGGGACACCTATGCTGAGTATGAGGCTGGTACAAGGC 1532  
Db 1422 AGGAA-----AGCAGAGGAGCGGTGCTCTTTGGAGAGATGGCGCTGGTAAAGC 1475  
QY 1533 CTGCAAGTGAAGCAGCCACAGAGTGAACACTACTCTGAGAAACCTGGGAGCTCTGTATAG 1592  
Db 1476 CTGCAAGTGAAGCAGCTCCACCGTCACAAACACCTTGAAACCTGGGAGCACTTACCG 1535  
QY 1593 GCGCAGGAGAAAGCTGGAGGCTGTGAGACCCCTGGAGAAATGTGCCCTGGGTCGCCGAG 1652  
Db 1536 CGCCAGGGGGAAGTTGAAGTGGGAGACACTGTGAGGAAGCTGCCCTGAGGTCAACGTA 1595  
QY 1653 ACAGGCACTGACCTATACGACAGCAAGGTGCGAGAGCTGCTGGGAGAGTGA 1709  
Db 1596 CGAGGCTTGACATGTTCAACACAGAGAGTGGCTGAAGTGCTAAATGACCCCTGA 1652

RESULT 6  
CGKINID CGKINID 2055 bp mRNA ROD 16-FEB-1998  
LOCUS Cricetus griseus mRNA for kinesin light chain isoform D.  
DEFINITION  
ACCESSION Y14586  
VERSION Y14586.1 GI:2330584  
KEYWORDS kinesin light chain.  
SOURCE Chinese hamster.  
ORGANISM Cricetus griseus  
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;  
Rodentia; Sciurognath; Muridae; Cricetinae; Cricetulus.  
REFERENCE 1 (Bases 1 to 2055)  
Khodjakov, A., Lizunova, E.M., Minin, A.A., Koonce, M.P. and  
Gyoeva, F.K.  
TITLE A specific light chain of kinesin associates with mitochondria in  
cultured cells  
JOURNAL Mol. Biol. Cell 9 (2), 333-343 (1998)  
MEDLINE 98119746  
REFERENCE 2 (bases 1 to 2055)  
Khodjakov, A.L.  
AUTHORS Direct Submission  
TITLE Submitted (13-AUG-1997) A.L. Khodjakov, Wadsworth Center, Empire  
JOURNAL State Plaza, PO Box 509, Albany, NY 12201-0509, USA  
FEATURES  
source location/Qualifiers  
1. .2055  
/organism="Cricetus griseus"  
/db\_xref="taxon:10029"  
/cell\_line="CHO-K1"  
/clone\_lib="Unikap (Stratagene)"  
1429. .1647  
/note="Protein sequence is in conflict with the conceptual  
translation"  
/codon\_start=1  
/product="kinesin light chain isoform D"  
/protein\_id="CAA74927.1"  
/db\_xref="GI:2330585"  
/db\_xref="SPTREMBL:O35500"  
/translation="VMSVWNGDGTGSLKRSFSKLSIRSRSEKLVRKLKGS  
RDSEPRNPASPAEPLCVENVQQPRR"

CDS

BASE COUNT 548 a 502 c 615 g 390 t  
ORIGIN

Query Match 31.1%; Score 762.8; DB 12; Length 2055;  
Best Local Similarity 69.9%; Pred. No. 7.4e-171;  
Matches 1076; Conservative 0; Mismatches 452; Indels 12; Gaps 3;

QY 356 TGTGTCATGAGAGAGCCGCGAGCTTCGCGCTTCTATGAAAAAATTGAGCTCGGCTGA 415  
Db 20 TGTGAGGAGAGTCCAGCATGATCCCGAAGTCTTGGAGATGCTGGAGCTCGGCTGA 79  
QY 416 GTGAGCCAGGTGATGCTGGCTTAGCCAGCACACTGAGCACAGTGGAGTCEGAGAAC 475  
Db 80 GTGAGCGCAGGTGATGATGGCCCTGTCCAATCACCTGAATCCCTGGAGTCCGAGAAC 139  
QY 476 AGAAGCTGGCGGCTCAGGTGCGCGGCTATGCCAGGAGACCACTGGCTCGCGGATGAGC 535  
Db 140 AGAAGCTGGCGGCTCAGGTGCGCGGCTGTGCCAGGAGAACAGTGGCTTCGCGGATGAAC 199  
QY 536 TGTGTCGACCCAGCAGCGGCTACAGCGCAGTGAACAGGCTGTGGCTCAGCTGGAGGAGG 595  
Db 200 TGSCCAACACGCGCAGAGAGTTGCAGAGAGTGAACAGTCCGCTGAGCTGAGGAGAG 259  
QY 596 AAAAGAACACCTGGAGTTCCTGGGCGAGCTGCGGCGAGTATGATGAGGATGGACATACCT 655  
Db 260 AGAAGAACACCTGGAGTTCATGAACACAGCTGAAGAAATATGACGATGACATCTCTCCCT 319  
QY 656 CGGAGGAGAGAGAGCGGATGCCACCAAGGATTCCTGGATGACCTCTTCCCTAATGAGG 715  
Db 320 CGGAGGAGAGAGACTCTGATTTACCAAGAGCCATTTGATGACCTCTTCCCAATGAGC 379  
QY 716 AGAAGA---GGACCCAGCAATGGCTTCTCCGTTGCTCAAGGTGCTACAGCAGCTCAGC 772  
Db 380 AGGATGACCCAGACAGGAATCCAGCAGACACAGCAGCGCTGGCGCCGCCAGC 439  
QY 773 AGGTTGATATGATATCCAGCAAGTTGCGGACGTTGCGAACCTGTGTGATCCAGTACG 832  
Db 440 AGGGGGCTACGAGATCCCGGCGCAGCTGCGCAGCGCTGCACAACTGGTATCCAGTATG 499  
QY 833 CAGCCCAAGTGCATGAGTGGCGGCTGCTGCTGTAAGCAGCAGCTAGAGGACCTGG 892  
Db 500 CTTACAGGGGCTTACGAGGTGGCGGTCCTCTGCAAGCAGCGCTTGGAGGACCTGG 559  
QY 893 AGCGACATCAGGCGGTCGCCACCTGTATGTCGCCACCATGCTCAACATCTCTGCTTTGG 952  
Db 560 AGAAGATTCGCGGCAATGACCACTGATGTGGCCACCATGCTGACATCTCTGGCCCTGG 619  
QY 953 TGTATGTCAGCAGATAAGTATAGGAAGCTGCCCCACCTGCTGTAATGATGCCCTTAGCA 1012  
Db 620 TGTACAGGATCAGAACAGTACAAAGATGCAAGCAATCTGCTGAACGACGCGCTGGCTA 679  
QY 1013 TCCGGGAGAGACCTTGGGACCTGACCATCTGCTGCTGGCTGCCACACTCAACATTTGG 1072  
Db 680 TCGTGAGAAACCTTGGCCGAGATCACCCCGGCTGCGCAGCAACTCTCAACACCTCG 739  
QY 1073 CTGTGCTCTATGCAAAAGGGCAAGTACAGAGGCGAGAGCTCTGTGCCAGCGGGCAC 1132  
Db 740 CAGTACTGTATGTAACGAGGAGTACAGAGGCGGACCGGCTGTGTAACAGGCT 799  
QY 1133 TGGAGTTGCAAAAGTCTTGGGCGCAGATCATCTCCAGATGTGCAAAACAGCTGAACA 1192  
Db 800 TGGAGATTAGAGAAGTTCTTGGGAAAGGATCATCCGAGTGTAGCCAAAGAGTTAAATA 859  
QY 1193 ACTGTGCTCTTGTGCCAAACACAGGCAAGTATGAGCGCTGGAAAGCTACTACGAGC 1252  
Db 860 ACTGTGCTCTTGTGCCAAACACAGGCAAGTATGAGGAGGTGGAGTATTACTACCA 919  
QY 1253 GAGCACTGCGCCATCTACGAGGCGGCTGGGCGCGGCAACCTTAATGTAGCCGAGCA 1312  
Db 920 GGCGCTGAGATTATCCAGACAAAGCTAGGACCGCGATGATCCCAATGTGGCCAAAGCA 979  
QY 1313 AGAACAACTGGCTTCTCTTACCTGAAACAGGCGCAATATGCTGAGGCTGAGACACTAT 1372



Db	864	ACATCTGGCCCTGGTGTACAGGATCAGAACAGTATAAAGATCAGCTAACCTCTCTGA	923
Qy	998	ATGATGCTTACATCCGGGAGACACCTTGGGACCTCACCATCCTCTCTGGCTGCCA	1057
Db	924	ACGACCCCTGGCTATCCGGAGAAACCTCGGCAGAGATCACCCTGGTGGCAGGA	983
Qy	1058	CACCTAACATTTGGTGTCTGTATGGGAAAGGGGCAAGTACAGAGGACAGCCCT	1117
Db	984	CTCTGACACCTAGCAGTACTGTAGCTTAAGCAGGAGAACTACAGGAGCGGACGC	1043
Qy	1118	TGTCAGGCGGACCTGGAGATTCGAGAAAGTCTCTGGCCAGCAATCATCCAGATGG	1177
Db	1044	TGTGTAAACGAGCCCTGGAGATCAGGAGAAAGTCTGGGAAAGATCATCTGATGTG	1103
Qy	1178	CAAAACAGCTGAACACCTGGCCCTTGTGCAAAACAGGCGCAAGTATGAGGCGGTG	1237
Db	1104	CAAACAGTTAAATAACCTGGCCCTCTCTGCCAAGCAGGAGTAACTACGAGAGGTG	1163
Qy	1238	AACGCTACTACAGCAGCACTGGCCATCTACGAGGGGAGCTGGGGCGGA---CAACC	1294
Db	1164	AGTATTATACAGAGGCGCTGGGATCTACACAGCAAGCTGGGCGCGATCGTACTC	1223
Qy	1295	CTAATGTAGCCCGGCAAGAACACCTGGCTTCTGTACCTGAACAGGCAATATG	1354
Db	1224	CCAACTGGCCAAAGACCAAGAACACCTGGCTTCTGTATCTGAACAAAGGAAGTCA	1283
Qy	1355	CTGAGCTGAGACACTATCAAGAGATCTTGACCGGTGCCATCTACAGGAGTTGGT	1414
Db	1284	AGAGGAGAAACGCTGTACAGGAGATCTCACCCGCGCACCAAGCGGAGTTGGAT	1343
Qy	1415	CTGTGTATGATGACCAACCCCATCTGATGCATCAGAGGAGCGGAGAAATGAGCA	1474
Db	1344	CTGTGGACGAGAGACCAAGCCCATCTGGTGCACCTCAAGAGAGAGGA-----GT	1397
Qy	1475	AAAGCCGGCACCATGAGGTGGGACACCTATGCTGAGTATGGAGCTGGTACAGGCT	1534
Db	1398	GCAAAGCAAGCAAGAGAGCGGTGCGCTTTTGGAGAGTATGGCGCTGGTAAAGCCT	1457
Qy	1535	GCAAAGTGAGCAGCCCAAGTGAACACTACTCTGAGAACCTGGGAGCTCTGTATAGCC	1594
Db	1458	GCAAAGTGAGACCTCCACCGCTCACACCCCTTGAACACCTTGGAGCATTACCCAG	1517
Qy	1595	GCAAGGAAAGCTGGAGCTGTGAGACCTTGAGGAATGTGCCCTGGCGTCCCGGAGAC	1654
Db	1518	GCGAGGGAAGTTTGAAGCTGCAGAGACATTTGAAGAAGCCCATGAGGTACCTAAGC	1577
Qy	1655	AGGCACTGACCCCTATCAGCCAGCAGCAAGGTGGCAGAGCTGTTGGGAGAGTGTATG	1714
Db	1578	AGGCTGTGACATGTTTCAACACAGAGATGGCTGAGTGTAAATGACCTGAGAGCA	1637
RESULT	8		
LOCUS	HUMKINESLC	2308 bp	09-MAR-1994
DEFINITION	Homo sapiens kinesin light chain mRNA, complete cds.		
ACCESSION	L04733		
VERSION	L04733.1	GI:307084	
KEYWORDS	kinesin light chain.		
SOURCE	Homo sapiens (human).		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;		
AUTHORS	1 (bases 1 to 2308)		
TITLE	Cabeza-Arvelaiz, I., Shih, L.C., Hardman, N., Asselbergs, F., Bilbe, G.,		
JOURNAL	Schmitz, A., White, B., Siciliano, M.J. and Lachman, L.B.		
MEDLINE	Cloning and genetic characterization of the human kinesin		
FEATURES	light-chain (KLC) gene		
source	DNA Cell Biol. 12 (10), 881-892 (1993)		
	9409988		
	Location/Qualifiers		
	1..2308		
	/organism="Homo sapiens"		

BASE COUNT	618 a	545 c	705 g	440 t
ORIGIN				
Query Match	29.1%	Score 713.4;	DB 9;	Length 2308;
Best Local Similarity	68.7%	Pred. No. 4e-159;		
Matches 1029;	Conservative	0;	Mismatches 456;	Indels 12; Gaps 3;
Qy	216	GCTCAGCCAGAGAGAGATCCTGGGAGACACAGCGGTGTCAGCCAGGCTAGAGGCGCT	275	
Db	318	GCTTACACAGGATGAAATATTCTTAAGACAAAGCAAGTAATTCAGGCGCTTGAAGCTTT	377	
Qy	276	ACGCACTGACACACAGGCGCTGCTGCAAGCCTGTCACAGACCACTGAGTCTGTCAGCA	335	
Db	378	GAAGAATGAGACAAATCTCCATTTTACAAAGTTTCTGGAGACACTGAAGTGTTTGAAGAA	437	
Qy	336	GGAGGCCATGAGGAGGCGCTGTCATGAGAAAGCCCGGAGCTTCCCTCTTATGGA	395	
Db	438	AGATG---ATGAAGTAATTTGTTGGAGGAGAAATCAACATGATCCGGAAGTCACTGA	494	
Qy	396	AAACATTTGCTGCGGCTGAGTGAGGCCCGAGTGTGCTGCTAGCCAGCCACCTGAG	455	
Db	495	GATGTTGGAGCTCGGCTGAGTGAGGCACAGTTATGATGGCTTTGTCAATCACTGAA	554	
Qy	456	CACAGTGGAGTCCGAGAAACAGAGCTCGGCGCTCAGTGGCGGCTATGCCAGGAGAA	515	
Db	555	TGCTGTGGAGTCCGAGAGCAAGAACTGGTGGCAGGTTCTGCTGTCGTCAGGAGAA	614	
Qy	516	CGAGTGGCTGGGATGAGCTGGTGGGACCCAGCAGCGGCTACAGCAGTGAACAGGC	575	
Db	615	TCAGTGGCTACGGGATGAATCGCCACACAGCAGCAAGAACTGCAGAGAGTGAAGCATC	674	
Qy	576	TGTGGCTCAGCTGGAGGAGAAAGAACCTCGGAGTTCTTGGGCGAGCTGCGGCGACTA	635	
Db	675	TGTGGCTCACTGGAGGAGAGAGAGCATCTGGAGTTTATGATCAGCTAAAAAATA	734	
Qy	636	TGATGAGGATGGACATACCTCGGAGGAGAAAGCGGATGCCACCAAGGATTCCTCGA	695	
Db	735	TGATGAGGACATTTCCCCATCCGAGGACAAAGACACTGATTTACCAAGAGCCTCTGA	794	
Qy	696	TGACCTTTTCTTAATGAGGAGGAGAA---GGACCCAGCAATGCTGTCTCCCGGTCA	752	
Db	795	TGACCTTTTCTTAATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG	854	
Qy	753	AGGTGTACAGCAGCTCAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG	812	
Db	855	TGCAGCGCGGCTGCCAGCAGGCGGCTACGAGATCCCGCGCGCTGCGGAGCCTCA	914	
Qy	813	CAACCTGGTATCCAGTACGACGACCCCAAGGTGCTATGAGTGGCGGTGCGGCTCTGTA	872	
Db	915	CAACCTGGTATCCAGTACGACGCTCGCAGGGGCGCTACGAGGTAGCTGTGCTCCCTTG	974	

/db\_xref="taxon:9606"  
 /cell\_type="T-cell"  
 /tissue\_type="blood"  
 /tissue\_lib="lambda-gtil, PHA-stimulated T-cell"  
 277..1986  
 /note="putative"  
 /codon\_start=1  
 /function="membrane-bounded organelles transport"  
 /product="kinesin light chain"  
 /protein\_id="AA16576.1"  
 /db\_xref="GI:307085"

CDS

AUTHORS	Wedaman,K.P., Knight,A.E., Kendrick-Jones,J. and Scholey,J.M.
TITLE	Sequences of sea urchin kinesin light chain isoforms
JOURNAL	J. Mol. Biol. 231 (1), 155-158 (1993)
MEDLINE	93267648
FEATURES	Location/Qualifiers 1..2601 /organism="Strongylocentrotus purpuratus" /db_xref="taxon:7668" /tissue_type="unfertilized egg" /feature_lib="lambda-zap" 132..2081 /codon_start=1 /product="kinesin light chain isoform 1" /protein_id="AAA03057.1" /db_xref="gi:161525"
CDS	/translation="MSGSKLTPNNSGGQNLNQEQIITGTRVIGKLGQLKNEHNDILNSYQSRLMKDDPTGSNLDVEKDIEKSLESLGLGEAKVMALGHHLNMVEDRQKDLRAEVRVLQENTWLRDELAATQKQLSEQNLADELVKYKLEYKLYNSTKKDYAEKTDLQAASDPLDLGFDPDDGGGADESYPQPQTSGSVSAAGGYETPLRLFTLNLIQASOSRYEVAVPLCKOALEDKTSYGHBDPVDVATMLNIALVYRDQNKYEKAENLHDALAJREKTLGDHPAATAANNLVLYGRKYKEAMELCLKRALREIKVLGNLHVAKOLANLLCONQKYEVEVYTORALEIYKELGPDPPNVAKTKNNLAARADHPVAKAETLYKQVLTTRAHERFGLSDADKNKPFWQAEEREKGFKNADNPGLDYGWGHAAKVDSSRSSTPTTTLKNGALYRHOGKYDAABILEECAMSRRLNALVRETVELLQDLSLTDYPRSEMAKERHRSSGTPHGSTESVYSYKTDGSENGGKLRSSGLSKLRSVRRSSSTKLKLIKRGRESDDDGGMKRASSMVLPSRGNDESTPAPLQISORGRVSHDNLSRRQSNF"
BASE COUNT	793 a 590 c 649 g 569 t
ORIGIN	Query Match 24.2%; Score 593.4; DB 33; Length 2601; Best Local Similarity 64.7%; Pred. No. 1.3e-130; Matches 977; Conservative 0; Mismatches 496; Indels 36; Gaps 5;
QY	223 CAAGAGGAGATCTGGGAGCACACCGCTGGTCAGCCAGGCGTAGAGGCCCTACGCAGT 282
DB	192 CAGGAGCAGATCATCACAGCACACGGAGGTGATCAAGGGTCTTGGAACAGCTCAAGAAT 251
QY	283 GAACACAGCGCTGTGCAAAGCCTGTCCCAGACATTGAGTGTCTGCAGCAGGAGGC 342
DB	252 GAACACAATCACATCTCCACAGCCTTACCAGAGCTCAAGATGCTCAAGAAGACACA 311
QY	343 CATGAGGAAGG --- CTGGTGCATGAGAGGCCCGCGACGCTTCCCGTTCTATGGAAC 399
DB	312 CCAGGTGACTCTAACCTGGTGGAGAGAGACTGATATCATCGAGAAGTCCCTTGAGAT 371
QY	400 ATTGAGCTCGGCTGAGTGAGGCCAGGTGATGCTGGCTCTAGCCAGCCACCTTGAGCACA 459
DB	372 CTGAGCTTGGTCTGGGTGAGGCTAAGTAATGATGGCATTTGGTCAACCATCTCAACATG 431
QY	460 GTGAGTGGAGAAACAGAGCTGCGGGCTCAGTGC GGCGGCTATGCCAGGAGAACACAG 519
DB	432 GTGAGGCCGAGAAACAGAGCTACGAGCTCAGGTTCCGAAGGTTAGTCCAGGAGATACC 491
QY	520 TGGCTCGGGATGAGCTGGCTGGCACCCAGCGGCTACAGCGCAGTGAACAGGCTGTG 579
DB	492 TGTTGAGAGATGAGCTGGCTGGCCCCAACAGAAAGTACAGACGAGTGAACAGATCTA 551
QY	580 GCTCAGCTGGAGGAGAAAAAGACCTCTGGAGTTCTTGGGGCAGCTGGCGCAGTATGAT 639
DB	552 GCTGATCTAGAGGTCAAATCAAGCACCTGGAGTATATGAATCAATCAAGAAGTATGAT 611
QY	640 GAGATGGACATACCTCGGAGGAGAAAGAGCGATGCCACCAAGGATTCCTCTGGATGAC 699
DB	612 GAAGATCGACACAGATGAGGAGGATCATCATCATGATCCCTTAGATCTTGGTTTCCCA 671
QY	700 CTCCTTTCCTAATGAGGAGGAGAGGCCCCAGCAATGGCTTTGTC --- GTGGTCAAGGT 756
DB	672 GAAGATGATGATGGTTCAGCAGCAGACGAGAGCTATCCCCAGCCCCAACAGGTAGTGT 731
QY	757 GCTACAGCAGCTCAGCAGGTGGATATGAGATCCACAGAGGTTGGGACGTTGACACAC 816

Db 732 TCAGTATCAGCAGCAGCAGGAGTTATGAGATCCCTGCCGTCTAGCTACTCTACACAA 791  
QY 817 CTGGTGATCAGTACGAGCCCAAGGTGCGTATGAGTGGCCGTGCCATCTGTGTAAAGCAG 876  
Db 792 CTGGTTATACAGTACGCGCTCACAGAGTGCATGAGTGGCTTACCACTCTGTAAACAA 851  
QY 877 GCACTAGAGCAGCTGGAGCCACATCAGGCGCGTGGCCACCTGATCTGCCACCATGCTC 936  
Db 852 GCTCTAGAGATCTGGAGAGACATCAGGTCAATGATCATCCAGATGTACAAACATGCTC 911  
QY 937 AACATCCTCTGTTGGTGTATCTGACCAAGATAAAGTATAAGAAAGCTGCCACCTGCTG 996  
Db 912 AACATCTTCCCTCGTATACAGGATCAAAACAAATACAAAGAGCTGTGTACCTGCTA 971  
QY 997 AATGATCCCTTAGATCCGGGAGAGCAGCTTGGACCTTGACCATCTCTGCTGTGGCTGCC 1056  
Db 972 CATGATGCTCTTGTATCAGGAGAAAGAGCTAGGACCAAGATCATCTCGAGTTGCAGCC 1031  
QY 1057 ACATCAACAAATTTGGCTGTCTATGGCAAAAGGGCAAGTACAAAGAGGACAGACCT 1116  
Db 1032 ACATTGAATACCTAGCAGTGTGTATGTTAGAGAGGCAAGTACAAAGAGGACCAACCT 1091  
QY 1117 CTGTGCCAGGCGCATGTGAGATTCAGAAAGGCTCTGGCAGCAATCATCCAGATGTC 1176  
Db 1092 TTATGTAAGAGAGCCCTGGAGATCAGGAAAGGTTCTTGGGAAAGATCATCTGTATGA 1151  
QY 1177 GCAAAACAGCTGAACACCTGGCCCTTCTGTGCAAAACAGGCGCAAGTATGAGCCGTC 1236  
Db 1152 GCTAAGCACTAAATATCTGGCTCTACTCTGCCAAATCAAGGAAATATGAAGAGGTT 1211  
QY 1237 GAACGCTACTACAGGAGCAGCTGGCCATCTACAGAGGCGAGCTGGGCGCGGACACCT 1296  
Db 1212 GAGTGTACTACAGGAGCAGCTAGAAATCTATGAAAGAGCTTGGTCCCGATGATCT 1271  
QY 1297 AATGTAGCCGGACCAAGAACACCTGGTCTCTGTACCTGAAACAGGCGCAATATGCT 1356  
Db 1272 AACGTAGCAAGCAAGAAACAACTTGTGAGCTTACCTCAAGCAGGCAAGTACAG 1331  
QY 1357 GAGGCTGAGACATATACAAAGAGATCTTGACCCGTGCCCATGTACAGAGATTTGGT--- 1414  
Db 1332 GCTGACAGACCTCTACAGAGGTTCTTAACAGGCGCCACGAAAGGAGTTTGGTCTA 1391  
QY 1415 ----CTGTGATGATGACCAAGCCATCTGTGATGATGATGATGATGATGATGATGATG 1470  
Db 1392 AGCGCTGATGACAGGACAAACCTCTGTGATGATGATGATGATGATGATGATGATG 1451  
QY 1471 AGCAAAAGCGGACCATGAGGTTGGGACACCTATGCTGATGATGATGATGATGATGAT 1530  
Db 1452 GGTAAATC-----AAGGATACGACCGTATGTTGATGATGATGATGATGATGATG 1502  
QY 1531 GCTGCAAAAGT-----ACGACGCGCAGAGTGAACACTACTCTGAGAAAC 1575  
Db 1503 GCAGCTAAAGTGAATAGTGGTCTGCGATGATGATGATGATGATGATGATGATGATG 1562  
QY 1576 CTGGGAGCTCTGTATAGGCGCCAGGAAAGCTGGAGGCTGTGAGACCTCTGGAGAAAGT 1635  
Db 1563 CTTGGGCGCTCTACAGAGACAGAAAGTATGATGCTGAGAGATCTGAGAGAGTGT 1622  
QY 1636 GCCTCGGCTCCGGAGACAGGCACTGACCTATACGACAGCAAGGTGCGAGAGCTG 1695  
Db 1623 GCCATGAAGTACGCAAGAAATGCTTGGACATGTTGGTGGAGACCAAAAGTCCGTGAG 1682  
QY 1696 CTTGGGAG 1704  
Db 1683 CTAGGCCAG 1691

RESULT 10  
SUSKINLCIB 2685 bp mRNA INV 26-OCT-1993  
LOCUS Strongylocentrotus purpuratus kinesin light chain isoform 2 mRNA,  
DEFINITION complete cds.  
ACCESSION L10234

VERSTON L10234.1 GI:161527  
KEYWORDS isoform 2; kinesin light chain.  
SOURCE Strongylocentrotus purpuratus (library: lambda-ZAP) unfertilized  
egg cDNA to mRNA.  
ORGANISM Strongylocentrotus purpuratus  
Eukaryota; Metazoa; Echinodermata; Echinozoa; Echinoidae;  
Euechinozoa; Echinacea; Echinoida; Strongylocentrotidae;  
Strongylocentrotus.  
REFERENCE 1 (bases 1 to 2685)  
AUTHORS Wedaman,K.P., Knight,A.E., Kendrick-Jones,J. and Scholey,J.M.  
TITLE Sequences of sea urchin kinesin light chain isoforms  
JOURNAL J. Mol. Biol. 231 (1), 155-158 (1993)  
MEDLINE 93267648  
FEATURES  
Location/Qualifiers  
1..2685  
/organism="Strongylocentrotus purpuratus"  
/db\_xref="taxon:7668"  
/tissue\_type="unfertilized egg"  
/tissue\_lib="lambda-ZAP"  
132..2165  
/codon\_start=1  
/product="kinesin light chain isoform 2"  
/protein\_id="AAA03058.1"  
/db\_xref="GI:161528"  
/translation="MSGSKLSTPNNSGGGGLSQBIITGTREVIGLEQLKNEHND  
LNSLYOSLKKMKKDPGDSNVEEDTDIEKSLSELEGLGEAKYMAALGHLNHNVE  
AEQKLRQAVRRLVQENTWLRDLAATQKLOTSEONLADLEVKYKHLVYMSIKKYD  
EDRTPDEASDPLDLPEDDDGGADSEYPOPTGSGSYSAAGGYEIPARLTL  
HNLIQVASQSRVEVAVPLCKOALELXTHSGHDHVDVATMLNIALYVDONKYKEA  
GNLHLALAIKREKTLGPDHPAVATLNNLAVLYKRGYKEAPLCPKRALEIKREKVLG  
KDPDVALKNNLALICNOGKYEEVYIYQALEYIEKLGDDPNVAKTNLAA  
YLQGYKAAETLYKQVLRAREFEFLSADDDKNKPIWMAEEREKGFKNAPYG  
DYGGWKAATVDSRSSTPTVTTLLKNGALYRCKGYDAAEILECAMKSRNALDM  
VRETKVELLGDVLEIDPSEAMAKERHRRSSGTPRHGSTESYSEYKTDGSEKRR  
AKDRSRIPAGVVEIPRSPHVLVENGDCILRRSSLSKLRSVRSSTKLKLNKGR  
ESDDGGMKRRSSMVLPRGNDSTPAFIQLSQRGRVGHNDLSRRSGNF"

BASE COUNT 816 a 614 c 671 g 584 t  
ORIGIN

Query Match 24.2%; Score 593.4; DB 33; Length 2685;  
Best Local Similarity 64.7%; Pred. No. 1.3e-130;  
Matches 977; Conservative 0; Mismatches 496; Indels 36; Gaps 5;

QY 223 CAAGAGAGATCTCTGGGAGCAGCAGCGCTGGTCAGCAAGGCTAGAGGCCCTACGCAGT 282  
Db 192 CAGGAGCAGATCATCACAGGCACACGGAGGTGATCAAGGCTCTGNACAGCTCAAGAT 251  
QY 283 GAACACAGCGCGTGTGCAAGCCTGTCGCCAGACCATGAGTGTCTGCAGCAGGAGGC 342  
Db 252 GAACACAATGACATCTCTCAACAGCCTCTACCAGAGCTCAAGATGCTCAAGAAGACACA 311  
QY 343 CATCAGAGAGG---CTGGTGCATGAGAGCGCCCGCAGCTTCGCCGTTCTATGGAAAC 399  
Db 312 CCAGTCACTCTAACCTCTGGTGGAGAGAGCTGATATCATCGAGAGTCCCTTGAGAGT 371  
QY 400 ATTGAGCTCGGCTGAGTGGAGCCCGAGGTGATGCTGGCTTAGCCAGCCACCTGAGCACA 459  
Db 372 CTGAGCTTGTCTGGTGGAGGCTAAGGTAATGATGCGATGGGTCAACATCTCAACATG 431  
QY 460 GTGAGTTCGGAGAAACAGAGCTCGCGGTCTAGTGGCGGCTATGCCAGGAGAACAG 519  
Db 432 GTGGAGCCCGAGAAACAGAGCTACGAGCTCAGGTTTCGAGGTTAGTCCAGGAGAAATAC 491  
QY 520 TGGCTCGGGATGAGCTGGCTGGCACCAGCGCTACAGCGCAGTGNACAGGCTGTG 579  
Db 492 TGGTTGAGATGAGCTGGCTGGCTGCCACCAACAGAGTACGAGAGTGAAGAGATCTA 551  
QY 580 GCTCAGCTGGAGAGGAGAAAGACCTCTGGAGTCTCTGGGCGAGCTGGGCGAGTATGAT 639  
Db 552 GCTGATCTAGAGTCAAAATCAAGCAGCTGGAGTATATGAACCTCATCAAGAGTATGAT 611  
QY 640 GAGGATGACATACCTCGGAGAGAAAGAGCGGATGCCACCAAGGATTCCTCTGGATGAC 699







Db 212 GAACGGAAATCGCCGAGGTGACAGAGACAATGAAGAATCGGACATGCTGCGAAGAACAT 271  
Qy 393 GGAATAAATGAGCTCGGGCTCAGTGAGGCCAGGTGATGCTGGCTCTAGCGAGCCACT 452  
Db 272 CGAGAACATCGAATCGGCTCAGCGAGGCCAGGTGATGATGGCTTGACATCCCATCT 331  
Qy 453 GAGCAGTGGAGTCGGAGAAACAGAAAGCTGCGGCTCAGGTGCGGCGGTATGCCAGGA 512  
Db 332 GCAGAACATCGAGGCGGAGAACACAAGCTTAAGACACAGAGTGCAGCGCCCTCCACAGGA 391  
Qy 513 GAACAGTGGCTGCGGATGAGCTGGTGGACACAGCTGAGTGGGCGAGCTGCGGCA 572  
Db 392 GAACGCTGGCTGCGGAGAGCTGGCCAACTCAGCAAAAGTTCCAGGATCCGAGCA 451  
Qy 573 GGTGTGGCTGAGCTGGAGAGGAGAAAGAACCTGAGTTCCTGGGCGAGCTGCGGCA 632  
Db 452 GCTGTGGCCCACTAGAGAGAGAGAGAGACCTGGAGTTCATGGCTCGGTGAAGAA 511  
Qy 633 GTATGATGAGGATGACATACCTCGAGAGAGAGAAAGAGCGGATGCCACAAGATTCCT 692  
Db 512 GTACGACGAGAAAT---CAGGAGCAGGACGACCTTGGACAAAGTCCCGCAGCTCCCGT 568  
Qy 693 GGATGACCTCTTCTTAATGAGGAGAGAGAGACCCAGCAATGCTGCTCCGTTGCA 752  
Db 569 GGTGAGCTGTTCCCGGAGGAGAAACAGGACAGACACACATGTCGCCACTCCGCC 628  
Qy 753 AGGTGCTACAGAGCTCAGCAGGCTGATGATGAGATCCAGCAAGGTTCCGACGTTGCA 812  
Db 629 CAGCAGTTGCGCAACCACTCCGGCTACGAGATCCAGCGCTCCGCTACTCTGCA 688  
Qy 813 CAACCTGGTATCCAGTACGAGCCCAAGTGCCTGATGAGTGGCGGTGCGCCTCTGAA 872  
Db 689 CAACCTGGTCAATCAGTATGCTATCGCAGGAAAGATACAGAGTGGCGTTCCCTCGAA 748  
Qy 873 GCAGGCACTAGGAGCTGAGCGGACATCAGCGCTGCGCAGCTGATGTCGCCACCAT 932  
Db 749 GCAGGCACTAGGAGCTGCTGAAAGGCAAGTGTGTCAGGACCATCCGATGTAGCAACAT 808  
Qy 933 GCTCAACATCTTGTGCTGCTGATGCTGACCAAGATAGTAAGAGAGCTGCCACCT 992  
Db 809 GCTGAACATTTTGGCTCTCGTACCGCGATCAGAACAGTATAAGGAGCGCGCAATT 868  
Qy 993 GCTGATGATGCTTATGATCGGAGAGACCTTGGACCTGACCATCTGCTGTGTC 1052  
Db 869 ACTCAAGAGCTCTGCTGATTCGAGGAAGACACTGGCGGAGAACATCCACGCTGTC 928  
Qy 1053 TGCACACTCAACAAATTTGGCTGCTCTATGCAAAAGGGCAAGTACAGGAGGCGAGA 1112  
Db 929 GGCAGCTTGAACAACTTGGCGCTCTCTATGCAAAAGCTGCAAGTACAAAGATGCCGA 988  
Qy 1113 GCCTCTGTGCGGAGCGCTGAGATTCGAGAAAGTCTGTTGGGACCAATCATCCAGA 1172  
Db 989 GCCACTGTGAAGCGTCTGATGATCGAGAGAGGCTCTGGGAAAGGATCATCCCGA 1048  
Qy 1173 TGTGGCAAAACACTGACAACTGCTGCTTGTGCCAAACAGGCGCAAGTATGAGGC 1232  
Db 1049 TGTGGCAAAACACTTAACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1108  
Qy 1233 CGTGGAAAGCTTACTACAGGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1292  
Db 1109 GGTGAGAACTACTACAGGAGCTTCTGCAATCTACGATCAAAATTTGGTCCCGATGA 1168  
Qy 1293 CCTATATGATGCGCGGACCAAGCAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1352  
Db 1169 TCCCAAGCTGCGCAAGCAAGCAAGTATCTGCGCGCTGCTGCTGCTGCTGCTGCTGCTG 1228  
Qy 1353 TGTGAGGCTGAGACACTATACAAAGAGATCTGACCGCTGCTGCTGCTGCTGCTGCTGCT 1412  
Db 1229 CACGAAAGCTGAAATCTCTATAGCAGGCTTGTAGCGGAGCCCAAGAGCTGAGTTGG 1288  
Qy 1413 GTCTGTGATGATGACACAGGCTTCTGATGATGATGATGATGATGATGATGATGATGATG 1472  
Db 1289 AGCATCGACAGCAAAACCAAGCCCAATTTGGCAGGTGGCTGAGGAGCGTGAGGA----- 1342

Qy 1473 CAAAAGCCGCGACCATGAGGTTGGGACACCCCTATGCTGAGTATGAGGCTGCTACAAGGC 1532  
Db 1343 GCACAAATTCGATAACAGGAGAGAACTCCATATGCGGATACGGCGTGGCATAAGGC 1402  
Qy 1533 CTCGAAAGTGGAGCGCCACAGTGAACACTACTCTGAGAAACCTGGGAGCTCTGATAG 1592  
Db 1403 CGTAAAGTGGATTCGCCCGGCTCAACCACTCTCTAAATAATCTGGGAGCACTTTACCG 1462  
Qy 1593 GCGCCAGGAAAGCTGGAGCTGCTGAGACCTGGAGGAATGTCCTGCGCTGCCGCGG 1652  
Db 1463 ACCTAAGCATGTTTGAAGCGCCGAAACCTGGAGACTGTCGAATGCGAGTAAATA 1522  
Qy 1653 ACAGGCACTGACCCCTATCAGCCAGAGCAAGTGGGAGAGCTGCT 1697  
Db 1523 AGAAGCTACGATCTAGCTAAACAAAGCAAGCTCTCACAATTGCT 1567

RESULT 13  
DROCKIN 1846 bp mRNA INV 26-APR-1993  
LOCUS Drosophila melanogaster (clone pBS-13a) kinesin light chain mRNA,  
DEFINITION complete cds.  
ACCESSION L11328  
VERSION L11328.1 GI:157813  
KEYWORDS kinesin light chain.  
SOURCE Drosophila melanogaster (library: lambda-gt11 of Salvatore) adult  
head cDNA to mRNA.  
ORGANISM Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
REFERENCE 1 (bases 1 to 1846)  
AUTHORS Gauger,A.K. and Goldstein,L.S.B.  
TITLE The Drosophila kinesin light chain: Proposed structure and  
interaction with kinesin heavy chain  
JOURNAL Unpublished (1993)  
FEATURES Location/Qualifiers  
source 1. .1846  
/organism="Drosophila melanogaster"  
/db\_xref="taxon:7227"  
/dev\_stage="adult"  
/tissue\_type="head"  
/tissue\_lib="lambda-gt11 of Salvatore"  
5'UTR 1. .119  
/gene="Klc"  
/db\_xref="FlyBase:FBgn0010235"  
gene 1. .1846  
/gene="Klc"  
/note="Kinesin light chain"  
/allele=""  
/db\_xref="FlyBase:FBgn0010235"  
CDS 120. .1646  
/gene="Klc"  
/codon\_start=1  
/db\_xref="FlyBase:FBgn0010235"  
/product="kinesin light chain"  
/protein\_id="AAA28669.1"  
/db\_xref="GI:157814"  
/translation="MTQMSQDEIITNTKTVLQGLEALRVEHVSIMNGIAEVQKNEKS  
DMLRKNIENIELGSEAQVMALTSHLQNIIEAKKKLTKQVRRLLHOENAWLRDELANT  
QKQFQSEQLVAQLEEEKKLEFEMASVKYDENQDQDADCKSRDTPVPELPDENE  
DRHNSPTTFPANOTGYEIPARLRLHNLVIQASOGRYEAVPICKQALELIER  
TSQHDHPDVTMLNIALYVROONKYKEANLLNDALSIRGTGLNGENHPAATLNNL  
AVLYGRGYKDAEPLEKRALREKVLGDKHPDKVAKQNLNALLCQNOGKDEVEKY  
YQALDIYESKLPDPPNVAKTKNNLAGYKQGRYEAELIKVLTTRAHREFGAI  
DSNKPFIWOVAEREERHKFDNRENTPYGEYGGWHKAAKVDSPVTTLKNLGLYRRO  
GMEFAETLEDCAMRSKKEYADLAKQTKLSQLLTSNEKRRSKAINEDLDFSEKNAP  
"  
1649. .1846  
/gene="Klc"  
/db\_xref="FlyBase:FBgn0010235"  
BASE COUNT 576 a 451 c 449 g 370 t



ORIGIN

Query Match 23.3%; Score 571.8; DB 33; Length 1846;  
Best Local Similarity 64.8%; Pred. No. 1.8e-125;  
Matches 884; Conservative 0; Mismatches 472; Indels 9; Gaps 2;

QY 333 GCAGGAGGCGATGAGGAGGCTGTGATGAGAGGCGCGGCGGCTTCAT 392  
DB 212 GACCGAATCCGAGGTGAGAGGAGCAATGAAAGTCGGACATGCTCGGAAAGCAAT 271  
QY 393 GGAAGCAATGAGCTGGGCTGAGTGAGGCGCCAGAGTATGCTGCTTACGAGCGACCT 452  
DB 272 CGAAGCAATGAACTGGGCTGAGGAGGCGCGGCTGATGATGGCGCTGACATCCCATCT 331  
QY 453 GAGCAGAGTGGAGTGGAGAAACAGAGCTGCGGGCTGAGGTGCGGCGCTATCCAGGA 512  
DB 332 GCAGAACATCGAGGCGCGAGAGGACAGCTTAAGACACAGGTGCGGCGCTCCACAGGA 391  
QY 513 GAACAGTGGCTGGGATGAGTGGCTGACACCGACAGCGCTACAGCGAGTGAACA 572  
DB 392 GAAGCGCTGGCTCGGACGAGTGGCTGACACCTGACGAAAGTTCAGGCGATCCGAGCA 451  
QY 573 GGCTGGCTGAGTGGAGAGGAAAGAGACCTGGAGTTCCTGGGCGAGCTGGCGCA 632  
DB 452 GCTGGTGGCGCACTAGAGGAGGAGAGAGACCTGGAGTTCATGGCGCTGGTGAAGAA 511  
QY 633 GTATGATGAGTGGATACCTCGGAGGAGAAAGAGGCGATGCCACCAAGGATTCCT 692  
DB 512 GTACAGAGAAAT---CAGAGAGAGGAGCGGTGCGAGCAAGTCCGACCGATCCCGT 568  
QY 693 GGATGACCTTTCTTAATGAGGAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 752  
DB 569 GGTGAGAGTGTCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 628  
QY 753 AGGTGCTACAGAGTGGATGATGATGATGATGATGATGATGATGATGATGATGATG 812  
DB 629 CAGCGAGTGGCGAAGAGTTCGGGCTGAGAGATGAGGAGGAGGAGGAGGAGGAGGAG 688  
QY 813 CAACCTGGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 872  
DB 689 CAACCTGGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 748  
QY 873 CGAGGAGTGGAGGAGTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 932  
DB 749 GCAAGCATGGAAGTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 808  
QY 933 GCTCAACATCTGCTTTGGTGTATGCTGACAGAGTAAAGTAAAGAGGAGTGGCGACCT 992  
DB 809 GCTGACATTTGGCTCTCGTACCGGATGACAGAGTAAAGAGGAGGAGGAGGAGGAG 868  
QY 993 GCTGAATGATGCCCTTAGCATCGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1052  
DB 869 ACTCAAGAGCTCTGTCGATTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 928  
QY 1053 TGCCACACTGACAAATTTGGCTGCTATGCGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1112  
DB 929 GGCACAGTGGAGCAATTTGGCGCTCTATGCGAAGGAGGAGGAGGAGGAGGAGGAGGAG 988  
QY 1113 GCCTCTGTGCCAGGCGGAGTGGAGATTCGAGAAAGGTCCTGGGAGGAGGAGGAGGAG 1172  
DB 989 GCCACTGTGAAGGCTGCTTGGAGATCCGAGAGAGGTCCTGGGAAAGGATCATCCGGA 1048  
QY 1173 TGTGCCAAACAGCTGACACAGCTGGCGCTTGTGCCAAACAGGAGGAGGAGGAGGAGG 1232  
DB 1049 TGTGCCAAACAGCTTACAACTCGCGCTGCTGCGAAGTACGAGGAGGAGGAGGAGGAG 1108  
QY 1233 CTTGGAAGGCTACTACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1292  
DB 1109 GGTGAGAGTACTACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1168  
QY 1293 CCTTAATGAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1352

DB 1169 TCCCAAGCTGGCAAGACAAAGATAATCTGCGCGCTGCTATCTGAAGCAAGGTAGATA 1238  
QY 1353 TGTGAGGCTGAGACACTATACAAAGAGATCTGACCCGTCGCCATGTACAGGAGTTGG 1412  
DB 1229 CACCGAAGCTGAATCTCTATAAGCAGGTCTTGAAGGAGGAGGAGGAGGAGGAGGAG 1288  
QY 1413 GTCTGTGGATGATGACCAAGCCCATCTGATGATGATGATGATGATGATGATGATGATG 1472  
DB 1289 AGCATCGACAGCAAAACAGCCCATTTGGCAGGTGGCTGAGGAGCGTGAGGA----- 1342  
QY 1473 CAAAGCGGACCATGAGGTGGGACACCTATGCTGATGATGATGATGATGATGATGATGAT 1532  
DB 1343 GCACAAATTCGATAACAGGAGGAGACACCTCATATGCGAGTACGCGGTTGCGATAAGC 1402  
QY 1533 CTCACAAAGTGGAGCGGCGGACAGTGAACACTACTCTGAGAAACCTGGGAGCTCTATAG 1592  
DB 1403 CGTAAAGTGGATTCGCCACCGGTACACCACTCTAAATAATCTGGGAGCATTACCG 1462  
QY 1593 GCGCCAGGAAAGCTGGAGGCTGTGAGACCTGTGAGGAGGATGTCCTGCGGTCGCCGAG 1652  
DB 1463 ACGTCAAGGATGTTTGAAGCGGCGGAAACCTGGAAGACTGTGCAATGCGGAGTAAAA 1522  
QY 1553 ACAGGCACTGACCCCTATCAGCAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1697  
DB 1523 AGAAGCTAGCTAGCTAGCTAAACAAACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1567

RESULT 14  
AF055298  
LOCUS AF055298 9915 bp DNA circular SYN 08-FEB-1999  
DEFINITION Expression vector pPK121, complete sequence.  
ACCESSION AF055298  
VERSION AF055298.1 GI:3023154  
KEYWORDS  
SOURCE Expression vector pPK121.  
ORGANISM Expression vector pPK121  
REFERENCE 1 (bases 1 to 9915)  
AUTHORS Coy,D.L., Wagenbach,M. and Howard,J.  
TITLE Kinesin takes one 8-nm step for each ATP that it hydrolyzes  
JOURNAL J. Biol. Chem. 274 (6), 3667-3671 (1999)  
MEDLINE 99121108  
REFERENCE 2 (bases 1 to 9915)  
AUTHORS Coy,D.L., Wagenbach,M. and Howard,J.  
TITLE Mechanochemical Coupling Stoichiometry of Single Kinesin Motors  
JOURNAL Unpublished  
REFERENCE 3 (bases 1 to 9915)  
AUTHORS Coy,D.L., Wagenbach,M. and Howard,J.  
TITLE Direct Submission  
JOURNAL Submitted (23-MAR-1998) Physiology and Biophysics, University of Washington, Campus Box 357290, Seattle, WA 98195-7290, USA

FEATURES  
source  
1. .9915  
/organism="Expression vector pPK121"  
/db\_xref="taxon:74939"  
/note="for expression of Drosophila melanogaster kinesin heavy chain and light chain in Escherichia coli"  
complement(3648..4730)  
/gene="laci"  
/gene="laci"  
/note="Laci"  
/codon\_start=1  
/transl\_table=11  
/product="lac repressor"  
/protein\_id="AAD13356.1"  
/db\_xref="GI:3023158"

gene  
CDS  
complement(3648..4730)  
/gene="laci"  
/note="Laci"  
/codon\_start=1  
/transl\_table=11  
/product="lac repressor"  
/protein\_id="AAD13356.1"  
/db\_xref="GI:3023158"

translation="MKPTVLYDAEYAGVSYQTSRVYVQASHVSAKTRKVEAMAE  
LNYIPNRVAQLAGKSLGLIGVATSSLAHAPSOIVAIAKSRADOLGASVVSVERS  
GYEACAAVHLLAQRVSLIINYPIDODATAVEACTNVPALFLDYSDQTPINSII  
FSHEDTRGLVEHLVAGHQIALLAGPLSSVARSRLRAGHKILTRNQIPRIAREG  
DWSAMSGFOOTMQLNEGIVPTAMLVANDQMALGARAITESGLRVGADISVVGDDT  
EDSSCYIPPLTTIKQDFRLLGTSVDRLLQLSQQAVKGNOLLPSLVKRTTLAPNT







GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 15, 2000, 01:35:28 ; Search time 156 Seconds  
(without alignments)  
3934.110 Million cell updates/sec

Title: US-09-036-614A-2

Perfect score: 2453

Sequence: 1 GTGAAGTGGTGAAGAGGG.....AGATCAGTCTTNTNTNNG 2453

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 311585 seqs, 125096042 residues

Total number of hits satisfying chosen parameters: 623170

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_36.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	262	10.7	268	1	Human secreted pro
2	229.4	9.4	811	1	Gastric cancer ass
3	205.6	8.4	717	1	Gastric cancer ass
4	164.2	6.7	361	1	Human brain expres
5	155.6	6.3	354	1	Human brain expres
6	143.4	5.8	356	1	EST clone BK455. N
7	138	5.6	763	1	Gastric cancer ass
8	94.8	3.9	178	1	EST clone CF89. Ne
9	77.8	3.2	799	1	Nucleotide sequenc
10	77.8	3.2	9600	1	Vector plasmid pCM
11	77.8	3.2	10596	1	Plasmid pcisEBON f
12	77.8	3.2	10596	1	Plasmid pcisEBON f
13	72.8	3.2	10596	1	Nucleotide sequenc
14	72.8	3.0	32207	1	KSHV LUR DNA (nucl
15	72.8	3.0	137507	1	KSHV long unique c
16	65.4	2.7	795	1	FLGA insert stabil
17	52.4	2.1	2000	1	Gene encoding Plas
18	51.4	2.1	203	1	SCA2 gene CAG repe
19	51.4	2.1	203	1	Glutamine rich reg
20	50.8	2.1	234	1	Spino cerebellar at
21	49.6	2.0	2744	1	MisSP1-containing p
22	48	2.0	195	1	Spino cerebellar at
23	47.2	1.9	165	1	Glutamine rich reg
24	47.2	1.9	2004	1	Nephila clavipes s
25	47	1.9	2214	1	New DNA sequence i
26	47	1.9	3331	1	New DNA sequence i
27	45	1.8	1150	1	PfHRP-II. New vacc
28	45	1.8	1150	1	Histidine-rich pro
29	44.8	1.8	543	1	Antigen tc-7a gene
30	44.6	1.8	1179	1	Plasmodium falci pa
31	44.4	1.8	1209	1	Porcine acylglucos
32	43.4	1.8	1542	1	DNA encoding antig
33	42.8	1.7	316	1	Antigen tc-26h gen

34	42.6	1.7	2301	1	V20445
35	42.4	1.7	4697	1	V32839
36	42	1.7	2818	1	T10554
37	42	1.7	3376	1	Q75166
38	41.8	1.7	2338	1	Q14183
39	41.8	1.7	2338	1	V23249
40	41.8	1.7	2594	1	X07075
41	41.4	1.7	171	1	Q84834
42	41.2	1.7	154	1	Q84835
43	41.2	1.7	2385	1	T33590
44	41.2	1.7	2462	1	V35473
45	41.2	1.7	3120	1	V71029

ALIGNMENTS

RESULT 1

X39587  
ID X39587 standard; DNA; 268 BP.  
AC X39587;  
DT 21-JUN-1999 (first entry)  
DE Human secreted protein 5; EST SEQ ID No 185.  
KW Human; secreted protein; EST; expressed sequence tag; diagnosis;  
KW forensic; gene therapy; chromosome mapping; signal peptide;  
KW upstream regulatory sequence; cytokine activity; cell proliferation;  
KW differentiation; haematopoiesis regulation; tissue growth regulation;  
KW reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;  
KW thrombolytic; anti-inflammatory; tumour inhibition; ds.  
OS Homo sapiens.  
FN W09906551-A2.  
PD 11-FEB-1999.  
PF 31-JUL-1998; IBI235.  
PR 01-AUG-1997; US-905133.  
PA (GEST ) GENSET.  
PI Duclert A, Dumas Milne Edwards J, Lacroix B;  
DR WPI; 99-153781/13.  
DR P-PSDB; Y11521.  
PT New nucleic acids encoding human secreted - proteins obtained from  
PT cDNA libraries prepared from substantia nigra, cerebellum, surrenals  
PT and fetal brain tissue  
PS Claim 1; Page 319; 43app; English.  
CC X39440 to X39597 represent 5' expressed sequence tags (ESTs) for human  
CC secreted proteins, and encode the proteins given in Y11374 to Y11531,  
CC respectively. The proteins given represent the signal peptide and an  
CC N-terminal fragment of a secreted protein. The nucleic acid sequences  
CC can be used for producing secreted human gene products. They can also  
CC be used to develop products for diagnosis and therapy. The proteins  
CC obtained may have cytokine activity, cell proliferation/differentiation  
CC activity, haematopoiesis regulating activity, tissue growth regulating  
CC activity, reproductive hormone regulating activity, chemotactic/  
CC chemokinetic activity, haemostatic and thrombolytic activity, receptor/  
CC ligand activity, anti-inflammatory activity, tumour inhibition activity  
CC or other activities. The products can be used in forensic, gene therapy  
CC and chromosome mapping procedures. The sequences can also be used for  
CC obtaining corresponding promoter sequences. The nucleic acids encoding  
CC the signal peptide can be used for directing extracellular secretion of  
CC a polypeptide or the insertion of a polypeptide into a membrane, or  
CC importing a polypeptide into a cell.  
SQ Sequence 268 BP; 59 A; 63 C; 104 G; 40 T;

Query Match 10.7%; Score 262; DB 1; Length 268;  
Best Local Similarity 98.9%; Pred. No. 5.1e-57;  
Matches 262; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy	399	CATTGAGTCGGGTGAGTGAGGCCCGCCAGGTGATGCTGCTCTAGCCAGCCACCTGAGCAC	458
Dd	1	CATTGAGTCGGGTGAGTGAGGCCCGCCAGGTGATGCTGCTCTAGCCAGCCACCTGAGCAC	60
Qy	459	AGTGGAGTCGGAGAAACAGAGAGCTGCGGGTCTAGGTCCGGCGGGTATGCCAGGAGACCA	518
Dd	61	AGTGGAGTCGGAGAAACAGAGAGCTGCGGGTCTAGGTCCGGCGGGTATGCCAGGAGACCA	120

QY 519 GTGGCTCGGAGTGTGGTGGCACCACAGCGGCTACAGCGGAGTGAACAGGCTGT 578  
 DB 121 GTGGCTCGGAGTGTGGTGGCACCACAGCGGCTACAGCGGAGTGAACAGGCTGT 180  
 QY 579 GGCTCAGCTGGAGGAGAAAGAGACCTGGAGTTCCTCGGGCAGCTGGGAGTATGA 638  
 DB 181 GGCTCAGCTGGAGGAGAAAGAGACCTGGAGTTCCTCGGGCAGCTGGGAGTATGA 240  
 QY 639 TGAGGATGGACATACCTCGGAGGAG 663  
 DB 241 TGAGGATGGACATACCTCGGAGGAG 265

RESULT 2  
 X39709  
 ID X39709 standard; DNA; 811 BP.  
 AC X39709;  
 DE 02-JUL-1999 (first entry)  
 KW Gastric cancer associated gene.  
 KW Cancer associated antigen; diagnosis; research; treatment; human;  
 KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;  
 KW prostate cancer; ss.  
 OS Homo sapiens.  
 PN WO9904265-A2.  
 PD 28-JAN-1999.  
 PF 15-JUL-1998; U14679.  
 PR 22-JUN-1998; US-102322.  
 PR 17-JUL-1997; US-896164.  
 PR 10-OCT-1997; US-061599.  
 PR 10-OCT-1997; US-061765.  
 PR 10-OCT-1997; US-948705.  
 PR 11-OCT-1997; US-021697.  
 PA (LUDW-) LUDWIG INST CANCER RES.  
 PI Chen Y, Gout I, Gure A, Ohare M, Obata Y, Old LJ,  
 PI Pfeundschtuh M, Sahin U, Scanlan MJ, Stockert E,  
 PI Tureci O;  
 DR WPI; 99-132448/11.  
 PT New isolated cancer associated nucleic acids and polypeptides -  
 PT isolated using sera from cancer patients, used to develop products  
 PT for the diagnosis, monitoring or treatment of cancers  
 PS Claim 67; Page 518-519; 787pp; English.  
 CC The invention relates to a method for diagnosing a disorder characterised  
 CC by expression of a human cancer associated antigen precursor coded for by  
 CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a  
 CC biological sample isolated from a subject with an agent that specifically  
 CC binds to the NAM, an expression product or a fragment of an expression  
 CC product complexed with an HLA molecule; and (b) determining the  
 CC interaction between the agent and the NAM or the expression product as a  
 CC determination of the disorder. The products and methods can be used in  
 CC the diagnosis, monitoring, research, or treatment of conditions  
 CC characterised by the expression of various cancer associated antigens.  
 CC which are cancer associated antigen precursors expressed in human breast  
 CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and  
 CC lung cancer.  
 SQ Sequence 811 BP; 236 A; 181 C; 236 G; 151 T;

Query Match 9.4%; Score 229.4; DB 1; Length 811;  
 Best Local Similarity 64.0%; Pred. NO. 1.2e-48;  
 Matches 407; Conservative 0; Mismatches 221; Indels 8; Gaps 4;

QY 216 GCTCAGCAAGAGAGATCTCTGGGAGCACACGCGTGTGTCAGCCAGGGCTAGAGGCCCT 275  
 DB 111 GCTTACACAGGATGAATTTATTTCTAGACAAAGCAGTAATTCAGGGGCTGGAAGCTTT 170  
 QY 276 AGCAGTGAACACAGCGCTGTGTCGAAGCCGTGTCCAGACATTGATGTCTGCAGCA 335  
 DB 171 GAAGAATGAGCACAATTCCTATTTACAAAGTTTGTCTGGAGACACTGAAGTGTGGAAGAA 230  
 QY 336 GGGAGGCCATGAGGAGGGCTGGTGCATGAGAAGGCCCGCGCTCGCGTTCATGGA 395

DB 231 AGATG---ATCAAGTAATTTGGTGGAGGAGAAATCAACATGATCCGGAAGTCACCTGA 287  
 QY 396 AAACATTGAGCTCGGCTGAGTGGAGCCCAAGGTGATGCTGGCTTACCGCAGCCACCTTGAG 455  
 DB 288 GATGTTGGAGCTCGGCTGAGTGGAGGACAGGTTATGATGGCTTTGTCAATCACCTGAA 347  
 QY 456 CACAGTGGAGTCGGAGAAACAGAAAGCTGGGGCTCAGGTGCGGGGGTATGCCAGAGGAA 515  
 DB 348 TGCTGTGGAGTCGAGAGCAAGAACTCGGTGCGCAGGTTCTGCTGTGTCAGGAGAA 407  
 QY 516 CCAGTGGCTGGGATGAGCTGGCACCACAGCAGCGGTACAGCGCAGTGAACAGGC 575  
 DB 408 TCAGTGGCTACGGATGACTGGCCACACAGCAGCAAAATTCGAGAAGAGTGAAGCAGTC 467  
 QY 576 TGTGGCTCAGCTGGAGGAGAAAGAACCTGGAGTTCCTGGGGCAGCTCGGGCAGTA 635  
 DB 468 TGTGGCTCAACTGGAGGAGGAGAAAGCATCTGGAGTTTATGAATCAGCTAAAAAATA 527  
 QY 636 TGATGAGATGGACATACCTCGGAGGAGAAAGAGGCGATGCCACCAAGGATTCCTCGGA 695  
 DB 528 TGATGACACATTTCCCATCCGAGGACAAAGACACTGATTTACCAAGAGCGCTCTGGA 587  
 QY 696 TGACCTCTTCTTAATGAGGAGGAAGA---GGACCCCAAGCAATGGCTTGTCGGTGTCA 752  
 DB 588 TGACCTTTTCCCATGATGAGACAGCCAGGCGCAAGGAATCCAGCAGCAGCACAGCAG 647  
 QY 753 AGGTGCTACAGCTCAGCAGGCTGGATA-TGAGATCCCAGCAAGGTTGCGGACGTTGC 811  
 DB 648 TGCAGCGCGGCTGCCCAGCAAGCNGCTACNAAGATTCCCGCGCGCTCGGAGCGCTCC 707  
 QY 812 ACACCT-GGTGTATCCAGTACGCGAGCCCAAGGTGCG 846  
 DB 708 ACAACCTGGTGATTCAGTTCGCTCCTCNCANGGGGC 743

RESULT 3  
 X39711  
 ID X39711 standard; DNA; 717 BP.  
 AC X39711;  
 DT 02-JUL-1999 (first entry)  
 DE Gastric cancer associated gene.  
 KW Cancer associated antigen; diagnosis; research; treatment; human;  
 KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;  
 KW prostate cancer; ss.  
 OS Homo sapiens.  
 PN WO9904265-A2.  
 PD 28-JAN-1999.  
 PF 15-JUL-1998; U14679.  
 PR 22-JUN-1998; US-102322.  
 PR 17-JUL-1997; US-896164.  
 PR 10-OCT-1997; US-061599.  
 PR 10-OCT-1997; US-061765.  
 PR 10-OCT-1997; US-948705.  
 PR 11-OCT-1997; US-021697.  
 PA (LUDW-) LUDWIG INST CANCER RES.  
 PI Chen Y, Gout I, Gure A, Ohare M, Obata Y, Old LJ,  
 PI Pfeundschtuh M, Sahin U, Scanlan MJ, Stockert E,  
 PI Tureci O;  
 DR WPI; 99-132448/11.  
 PT New isolated cancer associated nucleic acids and polypeptides -  
 PT isolated using sera from cancer patients, used to develop products  
 PT for the diagnosis, monitoring or treatment of cancers  
 PS Claim 67; Page 519; 787pp; English.  
 CC The invention relates to a method for diagnosing a disorder characterised  
 CC by expression of a human cancer associated antigen precursor coded for by  
 CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a  
 CC biological sample isolated from a subject with an agent that specifically  
 CC binds to the NAM, an expression product or a fragment of an expression  
 CC product complexed with an HLA molecule; and (b) determining the  
 CC interaction between the agent and the NAM or the expression product as a  
 CC determination of the disorder. The products and methods can be used in  
 CC the diagnosis, monitoring, research, or treatment of conditions  
 CC characterised by the expression of various cancer associated antigens.  
 CC which are cancer associated antigen precursors expressed in human breast  
 CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and  
 CC lung cancer.







Query Match 5.6%; Score 138; DB 1; Length 763; Best Local Similarity 63.8%; Pred. No. 1.2e-25; Matches 224; Conservative 0; Mismatches 124; Indels 4; Gaps 1;									
QY	1358	AGGCTGAGACACTATACAAAGAGATCTGACCCGTCGCCATGTACAGGAGTTGGTCTG	1417						
Db	738	AGGCGGAACANTGTACAAAGNGATCTCACTTGTGCACATGAAGGGAGTTGGTCTG	679						
QY	1418	TGGATGATGACCAAGCCCATCTGGATGATGACAGAGGCGGGAGGAATGAGCAAA	1477						
Db	678	TAGATGTTGAATTAACCCATTTGGAATGAT-----TGTGAGGAAGAAGAATGCAA	623						
QY	1478	GCCGGCACCATAGGGTGGGACACCCCTATCTGAGTAGTGGAGGCTGTACAAGGCCCTGCA	1537						
Db	622	AGGAAGCAAAAGGATGGGACATCTTTTGGAGAGTATGCGGTGGTCAAAAGCNTGCA	563						
QY	1538	AAGTGACGACGCCACAGTGAACACTACTCTGAGAAACCTGGGAGTCTGTATAGCGGCC	1597						
Db	562	AAGTTGATAGNCCAACTGTTTACAACCACTCTAAAAACCTTGGGGCACITTTACAGACGTC	503						
QY	1598	AGGGAAGCTGGAGGCTGCTGAGACCTGGAGGAATGTCCTGCGTCCGCGGAGACAGG	1657						
Db	502	AAGCAATTTGAAGCTGCGAAGACGTTAGAAGAGCTGCTATGAGGCTCTCGTAAACAGG	443						
QY	1658	GCACTGACCTATCAGCCAGCAGCAAGTGCGAGAGTCTGTTGGGGAGAGTGA	1709						
Db	442	GTCTTGACAATGTTCAACAACAGAGGTGCGAGAAGTGCTCAATGACCCCTGA	391						
RESULT 8									
V87408/c standard; cDNA; 178 BP.									
ID	V87408								
AC	V87408								
DT	27-APR-1999	(first entry)							
DE	EST clone CF89.								
KW	Expressed sequence tag; secreted protein; haematopoiesis regulator;								
KW	tissue growth; activin; inhibin; tumour invasion suppressor; EST; human;								
KW	chemotaxis; chemokinesis; haemostasis; gene therapy; thrombolysis;								
KW	receptor; ligand; anti-inflammatory; tumour inhibitor; ds.								
OS	Homo sapiens.								
PN	WO9845435-A2.								
PD	15-OCT-1998.								
PF	10-APR-1998; U06954.								
PR	10-APR-1997; US-835913								
PA	(GEM) GENETICS INST INC.								
PI	Agostino MJ, Jacobs K, Lavalie ER, McCoy JM, Merberg D,								
PI	Racie LA, Spaulding V, Treacy M;								
DR	WPI; 99-070076/06.								
PT	New polynucleotides encoding human secreted proteins - derived from								
PT	e.g. human blood, kidney, foetal lung, placenta, testes, brain,								
PT	ovary, pituitary, retina and colon cDNA libraries								
PS	Claim 1; Page 555; 633pp; English.								
CC	This sequence represents an expressed sequence tag (EST), and is a								
CC	polynucleotide of the invention. The polynucleotides of the invention are								
CC	all secreted EST sequences isolated from a variety of human tissue								
CC	sources. The EST sequences and proteins encoded by them are predicted to								
CC	have useful biological activities which would make them suitable for								
CC	treating, preventing or ameliorating medical conditions in humans and								
CC	animals, although no supporting data is given. Suggested activities								
CC	include nutritional activity, immune stimulating or suppressing activity,								
CC	haematopoiesis regulating activity, tissue growth activity,								
CC	activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic								
CC	and thrombolytic activity, receptor/ligand activity, anti-inflammatory								
CC	activity, cadherin/tumour invasion suppressor activity, tumour inhibition								
CC	therapy. The EST sequences are also stated to be useful for gene								
SQ	Sequence 178 BP; 36 A; 39 C; 37 G; 66 T;								
Query Match 3.9%; Score 94.8; DB 1; Length 178; Best Local Similarity 70.8%; Pred. No. 5.4e-15;									









CDS	90173..90643
FT	/product= glycoprotein X
FT	complement(93636..94127)
FT	/*tag= k
FT	/*tag= l
FT	/product= interferon regulatory factor 4
FT	complement(111931..112443)
FT	/*tag= m
FT	/product= capsid protein IV
FT	complement(123808..127296)
CDS	/*tag= n
FT	/product= immediate early protein
PN	W09804576-AI.
PD	05-FEB-1998.
PF	22-JUL-1997; U13346.
PR	29-NOV-1996; US-757669.
PR	25-JUL-1996; US-686243.
PR	25-JUL-1996; US-686349.
PR	25-JUL-1996; US-686350.
PR	25-JUL-1996; US-687233.
PR	25-JUL-1996; US-688814.
PR	05-SEP-1996; US-708678.
PR	10-OCT-1996; US-728323.
PR	13-NOV-1996; US-747887.
PR	13-NOV-1996; US-748640.
PA	(UYCO ) UNIV COLUMBIA NEW YORK
FI	Bohenzky RA, Chang Y, Edelman IS, Moore PS, Russo JJ;
DR	WPI; 98-130615/12.
PT	New nucleic acid encoding Kaposi's sarcoma associated herpes virus
PT	proteins - useful for, e.g. detecting levels of HHV8 in, and
PT	preparation of vaccines for treatment of, HIV patients
PS	Example 2: Page 135-203; 230pp; English.
CC	This sequence represents the long unique region and terminal repeat of
CC	the Kaposi's sarcoma-associated herpes virus (KSHV). KSHV is also known
CC	as human herpes virus 8 (HHV8). This sequence contains the DNAs of the
CC	nucleic acid encode KSHV polyptides selected from: (a) viral
CC	invention which encodes KSHV polyptides selected from: (a) viral
CC	macrophage inflammatory protein (MIP) II; (b) viral interleukin-6 (IL-6);
CC	(c) viral tnf alpha; (d) complement-binding protein; glycoproteins B, M or L;
CC	(e) capsid protein IV encoded by ORF65; and (e) immediate early protein
CC	encoded by ORF73. Labelled probes for the nucleic acid, proteins encoded
CC	by it, and antibodies (Ab) specific for the proteins are useful for
CC	detecting HHV8, specifically for diagnosis of Kaposi's sarcoma, in body
CC	fluids or tissue samples. HHV8 infections can be treated with antisease
CC	or triplex forming molecules or agents that bind specifically to the
CC	protein. Ab may be used for prophylaxis or treatment of HHV8 infection,
CC	while the protein can be used in protective vaccines. Ab may also be used
CC	to differentiate between lymphomas, and HHV8 may be implicated in many
CC	other lymphoproliferative diseases such as lymphomas, leukaemia,
CC	splenomegaly and mycosis fungoides. Cells and animals containing the
CC	nucleic acid are useful for drug screening. HHV8-derived peptides can be
CC	used as targets for antiviral drugs, e.g. dihydrofolate reductase gene
CC	can be inhibited with methotrexate. These can also be used to determine
CC	the immune status of a patient infected with HIV. HHV8 derived protein
CC	viral MIP III may be used as an anti-inflammatory agent for,
CC	e.g. treating rheumatoid arthritis. This sequence is stated as containing
CC	81 open reading frames.
SQ	Sequence 137507 BP; 32579 A; 37795 C; 35758 G; 31375 T;
Query Match 3.0%; Score 72.8; DB 1; Length 137507;	
Best Local Similarity 44.9%; Pred.No. 2.le-08;	
Matches 275; Conservative 0; Mismatches 337; Indels 0; Gaps 0	
QY	282 TGACACAGCGCCGCTGCTGC AAAGCTGTGCCAGCACCATTTGAGTGTCTGCAGCAGGGAGG 341 
Db	125323 TGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCA 125264 
QY	342 CCATGAGGAAGGCTGGTGATGAGNAGGCCCGGAGCTTCGCCGTTCATTGGAACAAT 401 
Db	125263 TGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCA 125204 
QY	402 TGAGCTCGGGCTGAGTGAGGCCAGGTGATGCTGGCTCTAGCCAGCACCTTGACACAGT 461 







GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 15, 2000, 00:40:42 ; Search time 52.1 Seconds  
(without alignments)  
6476.477 Million cell updates/sec

Title: US-09-036-614A-2  
Perfect score: 2453  
Sequence: 1 GTGAAGTGGTGAAGAGG.....AGATCAGTCTTTNTTNNNG 2453

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 243080 seqs, 6877915 residues

Total number of hits satisfying chosen parameters: 486160

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_NA.\*  
1: /cgn2.6/ptodata/1/ina/5A\_COMB.seq.\*  
2: /cgn2.6/ptodata/1/ina/5B\_COMB.seq.\*  
3: /cgn2.6/ptodata/1/ina/5C\_COMB.seq.\*  
4: /cgn2.6/ptodata/1/ina/5D\_COMB.seq.\*  
5: /cgn2.6/ptodata/1/ina/5E\_COMB.seq.\*  
6: /cgn2.6/ptodata/1/ina/PTUS\_COMB.seq.\*  
7: /cgn2.6/ptodata/1/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
c 1	77.8	3.2	5452	4 US-09-130-114-1	Sequence 1, Appl
2	77.8	3.2	10596	1 US-07-884-811-15	Sequence 15, Appl
3	77.8	3.2	10596	1 US-07-885-971-15	Sequence 15, Appl
4	77.8	3.2	10596	1 US-08-087-783A-15	Sequence 15, Appl
5	77.8	3.2	10596	1 US-08-194-088B-15	Sequence 15, Appl
6	77.8	3.2	10596	3 US-08-194-087-15	Sequence 15, Appl
7	77.8	3.2	10596	6 PCT-US93-04648-15	Sequence 15, Appl
8	72.8	3.0	3489	4 US-08-728-323A-1	Sequence 1, Appl
9	72.8	3.0	32207	3 US-08-770-379-20	Sequence 20, Appl
c 10	59.6	2.4	7218	1 US-08-232-463-14	Sequence 14, Appl
c 11	55.6	2.3	9551	4 US-08-056-200-93	Sequence 93, Appl
12	55.6	2.3	9551	4 US-08-800-644-93	Sequence 93, Appl
13	50.8	2.1	234	2 US-08-469-802B-3	Sequence 3, Appl
14	50.8	2.1	234	3 US-08-267-803B-3	Sequence 3, Appl
15	48	2.0	195	2 US-08-469-802B-2	Sequence 2, Appl
16	48	2.0	195	3 US-08-267-803B-2	Sequence 2, Appl
17	47	1.9	2214	5 US-08-864-038A-1	Sequence 1, Appl
18	47	1.9	3331	5 US-08-864-038A-2	Sequence 2, Appl
19	47	1.9	3331	5 US-08-864-038A-4	Sequence 4, Appl
20	46.4	1.9	2793	2 US-08-209-747-1	Sequence 1, Appl
21	46.4	1.9	2793	1 US-08-458-298-1	Sequence 1, Appl
22	45.8	1.9	7218	2 US-08-232-463-14	Sequence 14, Appl
c 23	45	1.8	1150	1 US-08-161-406-1	Sequence 1, Appl
24	44.8	1.8	543	7 5273901-6	Patent No. 5273901
25	44.4	1.8	1209	2 US-08-553-703A-5	Sequence 5, Appl
26	44.4	1.8	1209	2 US-08-553-703A-9	Sequence 9, Appl

ALIGNMENTS

RESULT 1

US-09-130-114-1/c  
; Sequence 1, Application US/09130114

; Patent No. 5976807

; GENERAL INFORMATION:

; APPLICANT: Horlick, Robert A.

; APPLICANT: Damaj, Bassam B.

; APPLICANT: Robbins, Alan K.

; TITLE OF INVENTION: Eukaryotic Cells Stably Expressing Genes

; TITLE OF INVENTION: From Multiple Transfected Episomes

; FILE REFERENCE: 0867/1D903US1

; CURRENT APPLICATION NUMBER: US/09/130,114

; CURRENT FILING DATE: 1998-08-06

; NUMBER OF SEQ ID NOS: 36

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 1

; LENGTH: 5452

; TYPE: DNA

; ORGANISM: VEBNA

US-09-130-114-1

Query Match 3.2%; Score 77.8; DB 4; Length 5452;

Best Local Similarity 47.6%; Pred. No. 1.4e-10;

Matches 263; Conservative 0; Mismatches 287; Indels 3; Gaps 1;

QY 175 GGCCTGTTGGGCGAGGGGATGAGCTTCAGGCCACCGGCTCAGCCCAAGAGGAGATC 234

DB 2117 GGGCAGGAGCAGGAGGAGGGGCGCAGGAGGAGGAGGAGGAGGAGGAGGAGG 2058

QY 235 CTGGGAGCACCGCTGGTTCAGCCAAAGGGCTAGAGGCCCTTACGCTGTAACACCGGCC 294

DB 2057 GAGGGCAGGAGCAGGAGGAGGGGCGCAGGAGGAGGAGGAGGAGGAGGAGGAGG 1998

QY 295 GTGCTGCAAAACCTGTCCACACCATTCAGTGTCTGCAGCAGGAGGCCATGAGGAGGG 354

DB 1997 CAGGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1938

QY 355 CTGGTGCATGAGAGGCCCGCCGCTTCGCGCTTCTATGAAAAACATTGAGCTCGGGCTG 414

DB 1937 GGGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1878

QY 415 AGTGAGGCCCGCAGGTGATCTGGCTCTAGCCAGCCACCTGAGCAGTGGAGTCGGAGAAA 474

DB 1877 GAGGGGAGGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1818

QY 475 CAGAAGCTGCGGGCTCAGTGGCGGGGCTATGCCAGGAGACCACTAGTGC---TCCGGGAT 531

DB 1817 CAGGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1758

QY 532 GAGCTGGCTGGACCCAGCAGCGGCTACAGCCAGTGAACAGGCTGTGGCTCAGCTGGAG 591  
 Db 1757 GAGGGCAGGAGGGGCAGGAGGGGCAGGAGCAGAGGGGGCAGGAGGGGGCAGGAG 1698  
 QY 592 GAGGAAAGAACACCTGAGTTCTGGGGCAGCTGCGGCAGTATGATGAGTGGACAT 651  
 Db 1697 CAGGAGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAG 1638  
 QY 652 ACCTCGAGAGGAAAGAGCGCATGCCAACAGGATTCCCTGGATGACCTCTTCTTAAT 711  
 Db 1637 GAGCAGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGGCAGGAGGGGCAGGAGGAG 1578  
 QY 712 GAGGAGGAAGAGG 724  
 Db 1577 GAGGGGCAGGAGG 1565

## RESULT 2

US-07-884-811-15

; Sequence 15, Application US/07884811

; Patent No. 5316921

; GENERAL INFORMATION:

; APPLICANT: Godowski, Paul J. Lokker, Nathalie A. Mark, Melanie R.

; TITLE OF INVENTION: SINGLE-CHAIN HEPATOCYTE GROWTH FACTOR VARIANTS

; NUMBER OF SEQUENCES: 21

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Genentech, Inc.

; STREET: 460 Point San Bruno Blvd

; CITY: South San Francisco

; STATE: California

; COUNTRY: USA

; ZIP: 94080

; COMPUTER READABLE FORM:

; MEDIUM TYPE: 5.25 inch, 360 kb floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: patin (Genentech)

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/07/884.811

; FILING DATE: 19920518

; CLASSIFICATION: 530

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER:

; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: Dreger, Ginger R.

; REGISTRATION NUMBER: 33,055

; REFERENCE/DOCKET NUMBER: 755.1

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 415/225-3216

; TELEFAX: 415/952-9881

; TELEX: 910/371-7168

; INFORMATION FOR SEQ ID NO: 15:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 10596 bases

; TYPE: NUCLEIC ACID

; STRANDEDNESS: single

; TOPOLOGY: linear

; US-07-884-811-15

## Query Match

Best Local Similarity 3.28; Score 77.8; DB 1; Length 10596;

Matches 263; Conservative 0; Mismatches 287; Indels 3; Gaps 1;

Matches 1;

QY 175 GCGCTGTGTTGGGCGACGGGATGAGCTGCAGGCCACCGCTCAGCCAAAGAGGATC 234

Db 2268 GGGCAGGAGCAGGAGGAGGGGCAGGAGGAGGGGCAGGAGGGGCAGGAGGGGCAG 2327

QY 235 CTGGGGACACACGGCTGTGTCAGCAAGGGGTAGGCCCTTACCGATGACACAGGCC 294

Db 2328 GAGGGGCAGGAGGAGGAGGGGCAGGAGGAGGGGCAGGAGGGGCAGGAGGGGCAG 2387

QY 295 GTGCTGCATAAGCCCTGTCTCCAGACCATTTAGTGTCTGCAGCAGGAGGCGCATGAGGAAGG 354  
 Db 2388 CAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGAGGAGGGGCAGGAGGAGGAGGAGGAG 2447  
 QY 355 CTGGTGCATGAGAAGCCCGCGGAGCTTCGCCGTTCTATGGAACAAATTGAGCTCGGGCTG 414  
 Db 2448 GGCAGGAGGGGCAGGAGGGGCAGGAGGAGGAGGGGCAGGAGGAGGAGGAGGAGGAG 2507  
 QY 415 AGTGAGGCCCCAGGTGATGCTGGCTCTAGCCAGCCACCTGACACAGTGGAGTCGGAGAAA 474  
 Db 2508 GAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGAGGGGCAGGAGGAGGGGCAGGAGGAG 2567  
 QY 475 CAGAAGCTCGGGCTCAGGTGCGGGCTTATGCCAGGAGAACCACTGGG ---TGGGGGAT 531  
 Db 2568 CAGGAGCAGGAGGGGCAGGAGGGGCAGGAGGAGGGGCAGGAGGAGGGGCAGGAGGAG 2627  
 QY 532 GAGCTGGCTGGCACCCAGCAGGCGGTACAGCCAGTGAACAGGCTGTGGCTCAGCTGGAG 591  
 Db 2628 GAGGGGCAGGAGGGGCAGGAGGGGCAGGAGGAGGGGCAGGAGGAGGGGCAGGAGGAG 2687  
 QY 592 GAGGAAAGAACACCTGGAGTTCTTGGGGCAGCTGCGGCAGTATGATGAGGATGGACAT 651  
 Db 2688 CAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGGAGGGGCAGGAGGGGCAG 2747  
 QY 652 ACCTCGGAGGAAAGAGCGCATGCCAACAGGATTCCCTGGATGACCTCTTTCTTAAT 711  
 Db 2748 GAGCAGGAGGGGCAGGAGGGGCAGGAGGAGGGGCAGGAGGAGGGGCAGGAGGAGGAG 2807  
 QY 712 GAGGAGGAAGAGG 724  
 Db 2808 GAGGGGCAGGAGG 2820

## RESULT 3

US-07-885-971-15

; Sequence 15, Application US/07885971

; Patent No. 5328837

; GENERAL INFORMATION:

; APPLICANT: Godowski, Paul J. Lokker, Nathalie A. Mark, Melanie R.

; TITLE OF INVENTION: HEPATOCYTE GROWTH FACTOR PROTEASE DOMAIN VARIANTS

; NUMBER OF SEQUENCES: 21

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Genentech, Inc.

; STREET: 460 Point San Bruno Blvd

; CITY: South San Francisco

; STATE: California

; COUNTRY: USA

; ZIP: 94080

; COMPUTER READABLE FORM:

; MEDIUM TYPE: 5.25 inch, 360 kb floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: patin (Genentech)

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/07/885.971

; FILING DATE: 19920518

; CLASSIFICATION: 530

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER:

; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: Dreger, Ginger R.

; REGISTRATION NUMBER: 33,055

; REFERENCE/DOCKET NUMBER: 779

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 415/225-3216

; TELEFAX: 415/952-9881

; TELEX: 910/371-7168

; INFORMATION FOR SEQ ID NO: 15:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 10596 bases

; TYPE: NUCLEIC ACID

STRANDEDNESS: single  
TOPOLOGY: linear  
US-07-885-971-15

Query Match 3.2%; Score 77.8; DB 1; Length 10596;  
Best Local Similarity 47.6%; Pred. No. 1.8e-10;  
Matches 263; Conservative 0; Mismatches 287; Indels 3; Gaps 1;

QY 175 GGCCTGGTGTGGGCGGAGTGGCTTCAGGCCACCGCTCAGCCAGAGGAGATC 234  
DB 2268 GGGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2327  
QY 235 CTGGGGAGCACACGGCTGGTTCAGCAAGGCTAGAGGCCCTACGACGTGAACACAGGCC 294  
DB 2328 GAGGGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2387  
QY 295 GTGCTGCAAGCCTGTCCAGACACATTGAGTGTCTGACGAGGAGGCCATGAGGAGGG 354  
DB 2388 CAGGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2447  
QY 355 CTGGTCATGAGAGGCCCGCGCAGCTTCGCGCTTCTATGGAACATTGAGCTCGGCTG 414  
DB 2448 GGGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2507  
QY 415 AGTGAGGCCAGGATGCTGCTTACGACACCTGACACAGTGGAGTGGAGAGAAA 474  
DB 2508 GAGGGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2567  
QY 475 CAGAGCTCGGGCTCAGTTCGGGGCTGTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAG 531  
DB 2568 CAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2627  
QY 532 GAGCTGGCTGGCACCAGCGGCTACAGCGAGTGAACAGGCTGTGGCTCAGCTGGAG 591  
DB 2628 GAGGGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2687  
QY 592 GAGGAAAGAACACCTGGAGTTCTTGGGGCAGCTTCGGGAGTATGATGAGGATGACAT 651  
DB 2688 CAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2747  
QY 652 ACCTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 711  
DB 2748 GAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2807  
QY 712 GAGGAGGAGGAGG 724  
DB 2808 GAGGGCAGGAGG 2820

RESULT 4  
US-08-087-783A-15  
Sequence 15, Application US/08087783A  
Patent No. 5547856  
GENERAL INFORMATION:  
APPLICANT: Godowski, Paul J., Lokker, Nathalie A., Mark, Melanie R.  
TITLE OF INVENTION: HEPATOCYTE GROWTH FACTOR VARIANTS  
NUMBER OF SEQUENCES: 22  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genentech, Inc.  
STREET: 460 Point San Bruno Blvd  
CITY: South San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94080  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: WinPatIn (Genentech)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/087,783A  
FILING DATE: 13-Jul-1993

CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/884811  
FILING DATE: 18-MAY-92  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/885971  
FILING DATE: 18-MAY-92  
ATTORNEY/AGENT INFORMATION:  
NAME: Marschang, Diane L.  
REGISTRATION NUMBER: 35,600  
REFERENCE/DOCKET NUMBER: P0755779P1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415/225-5416  
TELEFAX: 415/352-9881  
TELEX: 910/371-7168  
INFORMATION FOR SEQ ID NO: 15:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 10596 base pairs  
TYPE: Nucleic Acid  
STRANDEDNESS: Single  
TOPOLOGY: Linear  
US-08-087-783A-15

Query Match 3.2%; Score 77.8; DB 1; Length 10596;  
Best Local Similarity 47.6%; Pred. No. 1.8e-10;  
Matches 263; Conservative 0; Mismatches 287; Indels 3; Gaps 1;

QY 175 GGCCTGGTGTGGGCGGAGTGGCTTCAGGCCACCGCTCAGCCAGAGGAGATC 234  
DB 2268 GGGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2327  
QY 235 CTGGGGAGCACACGGCTGGTTCAGCAAGGCTAGAGGCCCTACGACGTGAACACAGGCC 294  
DB 2328 GAGGGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2387  
QY 295 GTGCTGCAAGCCTGTCCAGACACATTGAGTGTCTGACGAGGAGGCCATGAGGAGGG 354  
DB 2388 CAGGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2447  
QY 355 CTGGTCATGAGAGGCCCGCGCAGCTTCGCGCTTCTATGGAACATTGAGCTCGGCTG 414  
DB 2448 GGGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2507  
QY 415 AGTGAGGCCAGGATGCTGCTTACGACACCTGACACAGTGGAGTGGAGAGAAA 474  
DB 2508 GAGGGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2567  
QY 475 CAGAGCTCGGGCTCAGTTCGGGGCTGTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAG 531  
DB 2568 CAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2627  
QY 532 GAGCTGGCTGGCACCAGCGGCTACAGCGAGTGAACAGGCTGTGGCTCAGCTGGAG 591  
DB 2628 GAGGGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2687  
QY 592 GAGGAAAGAACACCTGGAGTTCTTGGGGCAGCTTCGGGAGTATGATGAGGATGACAT 651  
DB 2688 CAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2747  
QY 652 ACCTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 711  
DB 2748 GAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2807  
QY 712 GAGGAGGAGGAGG 724  
DB 2808 GAGGGCAGGAGG 2820

RESULT 5  
US-08-194-088B-15  
Sequence 15, Application US/08194088B  
Patent No. 5580963

GENERAL INFORMATION:  
APPLICANT: Godowski, Paul J. Lokker, Nathalie A. Mark, Melanie R.  
TITLE OF INVENTION: SINGLE-CHAIN HEPATOCYTE GROWTH FACTOR VARIANTS  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genentech, Inc.  
STREET: 460 Point San Bruno Blvd  
CITY: South San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94080  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: patin (Genentech)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/194,088B  
FILING DATE: 09-FEB-1994  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/884811  
FILING DATE: 18-MAY-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Gallegos, R. Thomas  
REGISTRATION NUMBER: 32,692  
REFERENCE/DOCKET NUMBER: 755D1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415/225-2614  
TELEFAX: 415/952-9881  
TELEX: 910/371-7168  
INFORMATION FOR SEQ ID NO: 15:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 10596 bases  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-194-088B-15

Query Match 3.2%; Score 77.8; DB 1; Length 10596;  
Best Local Similarity 47.6%; Pred. No. 1.8e-10;  
Matches 263; Conservative 0; Mismatches 287; Indels 3; Gaps 1;  
QY 175 GGCCTGGTGTGGGACGAGGATGCGCTGCAGCCACCGCTCAGCCACAGAGAGATC 234  
DB 268 GGGCAGGACGAGGAGGCGCAGGACGAGGAGGCGGCGAGGCGGCGAGGCGGCGAG 2327  
QY 235 CTGGGAGCACACGGCTGGTTCAGCCAGGCTTACGCGCTTACGCGTGAACACCGGCC 294  
DB 2328 GAGGCGCAGGACGAGGAGGCGGCGAGGAGGCGGCGAGGAGGCGGCGAGGAGGAG 2387  
QY 295 GTGCTGCAAGCCCTGCCACACCATTTAGTGTCTGCAGCAGGAGGCCATGAGGAAGG 354  
DB 2388 CAGGAGCAGGAGGAGGCGGCGAGGAGGCGGCGAGGAGGCGGCGAGGAGGAGGAG 2447  
QY 355 CTGGTGCATGAGAGGCGGCGGCGAGGCTTCCGCGTCTTATGGAAACATTGAGCTCGGCTG 414  
DB 2448 GGGCAGGAGGCGCAGGAGGCGGCGAGGACGAGGAGGCGGCGAGGAGGAGGCGGCGAG 2507  
QY 415 AGTGAAGCCCATGATGCTGGCTTACCGACCGCCCTGAGCAGCAGTGGAGTCGGAGAAA 474  
DB 2508 GAGGCGCAGGAGGAGGAGGCGGCGAGGAGGCGGCGAGGAGGCGGCGAGGAGGAGG 2567  
QY 475 CAGAAGCTGCGGGCTCAGGTGCGCGGCTATGCCAGGACACAGTGGC---TGCAGGAT 531  
DB 2568 CAGGAGCAGGAGGCGGCGAGGAGGCGGCGAGGAGGCGGCGAGGAGGCGGCGAGGAG 2627  
QY 532 GAGTGGCTGCGACCCAGCAGCGGCTACAGCGCAGTGAACAGGCTGTGGCTCAGCTGGAG 591  
DB 2628 GAGGCGCAGGAGGAGGCGGCGAGGAGGCGGCGAGGAGGCGGCGAGGAGGCGGCGAG 2687  
QY 592 GAGGAAAGAGACCTGGAGTTCCTGGGCGAGCTGGGCGAGCTGATGATGAGGATGACAT 651

DB 2688 CAGGAGGCGCAGGACGAGGCGGCGAGGAGGCGGCGAGGAGGCGGCGAGGAGGCGGCGAG 2747  
QY 652 ACCTCGGAGGAGAAAGCGGATGCCACCAAGGATTCCTGGATGACCTCTTTTCCTAAT 711  
DB 2748 GAGCAGGAGGCGGCGAGGAGGCGGCGAGGAGGCGGCGAGGAGGCGGCGAGGAGGCGGCGAG 2807  
QY 712 GAGGAGGAAGAGG 724  
DB 2808 GAGGCGCAGGAGG 2820  
RESULT 6  
US-08-194-087-15  
Sequence 15, Application US/08194087  
Patent No. 5879910  
GENERAL INFORMATION:  
APPLICANT: Godowski, Paul J. Lokker, Nathalie A. Mark, Melanie R.  
TITLE OF INVENTION: HEPATOCYTE GROWTH FACTOR PROTEASE DOMAIN VARIANTS  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genentech, Inc.  
STREET: 460 Point San Bruno Blvd  
CITY: South San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94080  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: patin (Genentech)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/194,087  
FILING DATE: 18-MAY-1992  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Dreger, Ginger R.  
REGISTRATION NUMBER: 33,055  
REFERENCE/DOCKET NUMBER: 779  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415/225-3216  
TELEFAX: 415/952-9881  
TELEX: 910/371-7168  
INFORMATION FOR SEQ ID NO: 15:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 10596 bases  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-194-087-15

Query Match 3.2%; Score 77.8; DB 3; Length 10596;  
Best Local Similarity 47.6%; Pred. No. 1.8e-10;  
Matches 263; Conservative 0; Mismatches 287; Indels 3; Gaps 1;  
QY 175 GGCCTGGTGTGGGACGAGGATGCGCTGCAGCCACCGCTCAGCCACAGAGAGATC 234  
DB 268 GGGCAGGACGAGGAGGCGGCGAGGAGGCGGCGAGGAGGCGGCGAGGAGGCGGCGAG 2327  
QY 235 CTGGGAGCACACGGCTGGTTCAGCCAGGCTTACGCGCTTACGCGTGAACACCGGCC 294  
DB 2328 GAGGCGCAGGACGAGGAGGCGGCGAGGAGGCGGCGAGGAGGCGGCGAGGAGGAG 2387  
QY 295 GTGCTGCAAGCCCTGCCACACCATTTAGTGTCTGCAGCAGGAGGCCATGAGGAAGG 354  
DB 2388 CAGGAGCAGGAGGAGGCGGCGAGGAGGCGGCGAGGAGGCGGCGAGGAGGAGGAG 2447  
QY 355 CTGGTGCATGAGAGGCGGCGGCGAGGCTTCCGCGTCTTATGGAAACATTGAGCTCGGCTG 414



```
; APPLICATION NUMBER: US/08/728, 323A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 0575/52268/JPW/MSC/SKS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-278-0400
; TELEFAX: 212-391-0525
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3489 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..3489
; US-08-728-323A-1

Query Match          3.0%; Score 72.8; DB 4; Length 3489;
Best Local Similarity 44.9%; Pred. No. 2.2e-09;
Matches 275; Conservative 0; Mismatches 337; Indels 0; Gaps 0;

QY 282 TGAACACCGCCGCTGCTGCAAAAGCCTGTCCAGACCAATGATGTCTCGACGAGGAG 341
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1974 TGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGGA 2033

QY 342 CCATGAGGAAGGCTGTGTCATCAGAAAGCCCGCGAGCTTCGCCGTTCTATGGAACAT 401
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2034 TGACACACACAGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCA 2093

QY 402 TGAGCTCGGGCTGAGTGAGGCCCGCAGGTGATGCTGGCTCTAGCCAGCCACTGAGCACAT 461
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2094 TGAGCAGGACGACGATGAGCAGCAGCAGCAGGATGAGCAGGACGACGAGGATGAGCA 2153

QY 462 GGATCGGAGAAACAGAGCTCGGGCTCAGGTGCGCGGCTATGCCAGGAGAACCATGTG 521
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2154 GCAGCAGGATGAGCAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCA 2213

QY 522 GCTGCGGGATGAGCTGCTGGCCACCCAGCAGCGGCTACAGCCAGTGAACAGCTGTGCG 581
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2214 GGATGACGACACAGGATGAGCAGCAGCAGCAGGATGAGCAGGACGACGAGGAGGA 2273

QY 582 TCAGCTGGAGGAGGAAAGAACCTTGAGTTCTTGGGCGAGCTCGCGCATATGATGA 641
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2274 GCAGGAGCAGCAGGAGGAGCAGGACGAGGATGAGGAGCAGGAGGAGGATGAGGAGGA 2333

QY 642 GGATGACATACCTCGAGGAGAAAGAGCGCATGCCACCAAGGATTCCTGGATGACCT 701
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2334 TCAGGAGCAGGAGGATGAGGAGCAGGAGGAGGATGAGGAGCAGGAGGAGGATGAGA 2393

QY 702 CTTTCTATGAGGAGGAGGAGGAGCCAGCAATGCTTGTCCCTGCTCAAGGTGCTTAC 761
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2394 GGAGCAGGACGAGGATGAGGAGCAGGAGCAGGAGGATGAGGAGCAGGAGGATGAGG 2453

QY 762 AGCAGCTCAGAGGGTGGATATGAGATCCAGCAAGGTTGCGGAGGTTGCAACACCTGGT 821
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2454 AGAGGAGCAGGAGCAGGAGTATGAGGAGCAGGAGGAGGATGAGGAGCAGGAGGAGGA 2513

QY 822 GATCCAGTACGCCCAAGGCTCGCTATGAGTGGCGCTGCCACTCTGTAGCAGGCACT 881
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2514 GTTAGGAGGAGCAGGAGGTGGAAGAGCAAGCAGGAGGATGGAAGAGCAAGCAGGAGCA 2573

QY 882 AGAGGACCTGGA 893
   ||| ||| ||| |||
Db 2574 GGAAGAGCAGGA 2585

RESULT 9
```

```
US-08-770-379-20/c
; Sequence 20, Application US/08770379
; Patent No. 5849564
; GENERAL INFORMATION:
; APPLICANT: Chang, Yuan
; APPLICANT: Bohenzky, Roy A.
; APPLICANT: Russo, James J.
; APPLICANT: Edelman, Isidore S.
; APPLICANT: Moore, Patrick S.
; TITLE OF INVENTION: POLYPEPTIDES FROM KAPOSI'S SARCOMA-ASSOCIATED
; HERPESVIRUS, DNA ENCODING SAME AND USES THEREOF
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/770,379
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 52342
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0525
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 32207 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-770-379-20
```

```
Query Match          3.0%; Score 72.8; DB 3; Length 32207;
Best Local Similarity 44.9%; Pred. No. 5.2e-09;
Matches 275; Conservative 0; Mismatches 337; Indels 0; Gaps 0;

QY 282 TGAACACCGCCGCTGCTGCAAAAGCCTGTCCAGACCAATGATGTCTCGACGAGGAGG 341
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2023 TGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGGA 19964

QY 342 CCATGAGGAAGGCTGTGTCATCAGAAAGCCCGCGAGCTTCGCCGTTCTATGGAACAT 401
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 19963 TGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGGA 19904

QY 402 TGAGCTCGGGCTGAGTGAGGCCCGCAGGTGATGCTGGCTCTAGCCAGCCACTGAGCACAT 461
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 19903 TGAGCAGGACGACGAGGATGAGCAGCAGCAGCAGGATGAGCAGGACGACGAGGATGAGCA 19844

QY 462 GGATCGGAGAAACAGAGCTGCGGGCTCAGGTGCGCGGCTATGCCAGGAGAACCATGTG 521
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 19843 GCAGCAGGATGAGCAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGCA 19784

QY 522 GCTGCGGGATGAGCTGCTGGCCACCCAGCAGCGGCTACAGCCAGTGAACAGCTGTGCG 581
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 19783 GGATGACGACGACGAGGATGAGCAGCAGCAGGATGAGCAGGAGGATGAGCAGCAGGAGGA 19724

QY 582 TCAGCTGGAGGAGGAAAGAACCACTTGGAGTTCTTGGGCGAGCTGCGGCAGTATGATGA 641
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 19723 GCAGGAGCAGCAGGAGGAGCAGGAGGATGAGGAGGAGGATGAGGAGGAGGAGGATGAGGA 19664
```



```
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1507..1644
; FEATURE:
; NAME/KEY: Intron
; LOCATION: 1645..2511
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2512..8070
US-08-056-200-93

Query Match      2.3%; Score 55.6; DB 1; Length 9551;
Best Local Similarity 46.4%; Pred. No. 7.9e-05;
Matches 181; Conservative 0; Mismatches 209; Indels 0; Gaps 0;

QY 281 GTGAACACAGGCGCTGCTGCAAGCGCTGCCAGACCATTTGAGTGTCTGCAGAGGGAG 340
Db 3356 GCGAGCAGCAGGAGAGAGCGCGAGCAGCTGAGGCGCGAGGAGGAGGAGCGCG 3415
QY 341 GCCATGAGGAGGCGCTGGTGCATGAGAAGCGCGCGAGCTTCGCCGCTTCTATGAAACA 400
Db 3416 AGCAGCAGCTGAGGCGCGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 3475
QY 401 TTGAGCTCGGGGTGAGTGGTGCATGAGAGCGCGCGAGCTTACGCCACCTAGCACAG 460
Db 3476 AGGAGGAGGAGGCGCGAGCAGCAGCTGAGGCGCGAGGAGGAGGAGGAGGAGGAGG 3535
QY 461 TGGAGTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 3535
Db 3476 AGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 3535
QY 461 TGGAGTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 520
Db 3536 AGCAGCTGAGGCGCGAGCAGCTGAGGCGCGAGCAGCTGAGGCGCGAGGAGGAGGAGG 3595
QY 521 GCCTCGGGGTGAGTGGTGCATGAGAAGCGCGCGAGCTTCGCCGCTTCTATGAAACA 580
Db 3596 TGAGGCGCGAGCAGCAGCTGAGGCGCGAGGAGGAGGAGGAGGAGGAGGAGGAGG 580
QY 581 CTCAGCTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 640
Db 3656 GCGAGCAGCAGCTGAGGCGCGAGCAGCTGAGGCGCGAGGAGGAGGAGGAGGAGGAGG 3715
QY 641 AGGATGACATACCTCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 670
Db 3716 AGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 3745

RESULT 12
US-08-800-644-93 Application US/08800644
; Sequence 93 Application US/08800644
; Patent No. 5958752
; GENERAL INFORMATION:
; APPLICANT: Steinert, Peter M.
; APPLICANT: Lee, Seung-Chul
; APPLICANT: Kim, In-Gyu
; APPLICANT: Chung, Soo-Il
; APPLICANT: Park, Sang-Chul
; TITLE OF INVENTION: Trichohyalin and Transglutaminase-3 and
; TITLE OF INVENTION: Methods of Using Same
; NUMBER OF SEQUENCES: 117
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe, Martens, Olson & Bear
; STREET: 620 Newport Center Drive, Sixteenth Floor
; CITY: Newport Beach
; STATE: CA
; COUNTRY: U.S.A.
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
```

```
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/800,644
; FILING DATE: 14-FEB-1997
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/056,200
; FILING DATE: 30-APR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Fedrick, Michael F.
; REGISTRATION NUMBER: 36,799
; REFERENCE/DOCKET NUMBER: NIH054.001A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (714) 760-0404
; TELEFAX: (714) 760-9502
; INFORMATION FOR SEQ ID NO: 93:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9551 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1507..1644
; FEATURE:
; NAME/KEY: Intron
; LOCATION: 1645..2511
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2512..8070
US-08-800-644-93

Query Match      2.3%; Score 55.6; DB 4; Length 9551;
Best Local Similarity 46.4%; Pred. No. 7.9e-05;
Matches 181; Conservative 0; Mismatches 209; Indels 0; Gaps 0;

QY 281 GTGAACACAGGCGCTGCTGCAAGCGCTGCCAGACCATTTGAGTGTCTGCAGAGGGAG 340
Db 3356 GCGAGCAGCAGGAGGAGAGCGCGAGCAGCTGAGGCGCGAGGAGGAGGAGGAGGAGG 3415
QY 341 GCCATGAGGAGGCGCTGGTGCATGAGAAGCGCGCGAGCTTCGCCGCTTCTATGAAACA 400
Db 3416 AGCAGCAGCTGAGGCGCGAGCAGCTGAGGCGCGAGGAGGAGGAGGAGGAGGAGGAGG 3475
QY 401 TTGAGCTCGGGGTGAGTGGTGCATGAGAGCGCGCGAGCTTACGCCACCTAGCACAG 460
Db 3476 AGGAGGAGGAGGCGCGAGCAGCAGCTGAGGCGCGAGGAGGAGGAGGAGGAGGAGGAGG 3535
QY 461 TGGAGTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 520
Db 3536 AGCAGCTGAGGCGCGAGCAGCTGAGGCGCGAGGAGGAGGAGGAGGAGGAGGAGGAGG 3595
QY 521 GCCTCGGGGTGAGTGGTGCATGAGAAGCGCGCGAGCTTCGCCGCTTCTATGAAACA 580
Db 3596 TGAGGCGCGAGCAGCAGCTGAGGCGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 580
QY 581 CTCAGCTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 640
Db 3656 GCGAGCAGCAGCTGAGGCGCGAGCAGCTGAGGCGCGAGGAGGAGGAGGAGGAGGAGG 3715
QY 641 AGGATGACATACCTCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 670
Db 3716 AGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 3745

RESULT 13
US-08-469-802B-3
; Sequence 3, Application US/08469802B
; Patent No. 5741645
; GENERAL INFORMATION:
```



RESULT 14	
US-08-267-803B-3	
; Sequence 3, Application US/08267803B	
; Patent No. 5834183	
; GENERAL INFORMATION:	
; APPLICANT: Orr, Harry T.	
; APPLICANT: Ranum, Laura P.W.	
; APPLICANT: Chung, Ming-yi	
; APPLICANT: Zoghbi, Huda Y.	
; TITLE OF INVENTION: Gene Sequence	for Spinocerebellar Ataxia
; Patent No. 5834183	

RECORD 15  
US-08-469-802B-2  
US-08-469-802B-2  
Sequence 2, Application US/08469802B  
Patent No. 5741645  
GENERAL INFORMATION:  
APPLICANT: Orr, Harry T.  
APPLICANT: Ranum, Laura P.W.  
APPLICANT: Chung, Ming-yi  
APPLICANT: Zoghbi, Huda Y.  
TITLE OF INVENTION: Gene Sequence for Spinocerebellar Ataxia  
Patent No. 5741645  
TITLE OF INVENTION: Type 1 and Method for Diagnosis  
NUMBER OF SEQUENCES: 47  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Mueiting, Raasch, Gebhardt & Schwappach, P.  
STREET: 119 No. 5741645th Fourth Street, Suite 203  
CITY: Minneapolis

```

: STATE: MN
: COUNTRY: USA
: ZIP: 55401
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent in Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/469,802B
: FILING DATE: 06-JUN-1995
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Muetting, Ann M.
: REGISTRATION NUMBER: 33,977
: REFERENCE/DOCKET NUMBER: 110.00030101
: TELECOMMUNICATION INFORMATION:
: * TELEPHONE: 612-305-1217
: TELEFAX: 612-305-1225
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 195 base pairs
: TYPE: nucleic acid
: * TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA
: US-08-469-802B-2

```

```

Query Match          2.0%; Score 48; DB 2; Length 195;
Best Local Similarity 55.4%; Pred. No. 0.0015;
Matches 93; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

Qy 468 GGAGAAACAGAAAGCTGCGGGCTATGCCAGGAGAACCAAGTGGCTGG 527
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3 GCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA 62
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 528 GGATGAGCTGGCTGGCACCAGCAGCGGCTACAGCGCAGTGAACAGCTGTGCTCAGCT 587
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 63 GCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA 122
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 588 GGAGGAGGAAAAAGAACACCTGGAGTTCTCTGGGGCAGCTGGCGAGTA 635
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 123 GCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA 170
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

Search completed: August 15, 2000, 04:31:12  
Job time: 13830 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 15, 2000, 00:19:33 ; Search time 875.55 Seconds  
(without alignments)  
12354.872 Million cell updates/sec

Title: US-09-036-614A-2  
Perfect score: 2453  
Sequence: 1 GTGAAGTGTGAAGAGGG.....AGAACTCACTTCTTNTINNG 2453

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 5247842 seqs, 2204914090 residues  
Total number of hits satisfying chosen parameters: 10495684

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
EST:  
1: em\_est1:  
2: em\_est2:  
3: em\_est3:  
4: em\_est4:  
5: em\_est5:  
6: em\_est6:  
7: em\_est7:  
8: em\_est8:  
9: em\_est9:  
10: em\_est10:  
11: em\_est11:  
12: em\_est12:  
13: em\_est13:  
14: em\_est14:  
15: em\_est15:  
16: em\_est16:  
17: em\_est17:  
18: em\_est18:  
19: em\_est19:  
20: gb\_est1:  
21: gb\_est2:  
22: gb\_est3:  
23: gb\_est4:  
24: gb\_est5:  
25: gb\_est6:  
26: gb\_est7:  
27: gb\_est8:  
28: gb\_est9:  
29: gb\_est10:  
30: gb\_est11:  
31: gb\_est12:  
32: gb\_est13:  
33: gb\_est14:  
34: gb\_est15:  
35: gb\_est16:  
36: gb\_est17:  
37: gb\_est18:  
38: gb\_est19:  
39: gb\_est20:  
40: gb\_est21:  
41: gb\_est22:  
42: gb\_est23:  
43: gb\_est24:  
44: gb\_est25:  
45: gb\_est26:  
46: gb\_est27:  
47: gb\_est28:  
48: gb\_est29:  
49: gb\_est30:  
50: gb\_est31:  
51: gb\_est32:  
52: em\_est20:  
53: em\_est21:  
54: em\_est22:  
55: em\_est23:  
56: em\_est24:  
57: em\_est25:  
58: em\_est26:  
59: gb\_est33:  
60: gb\_est34:  
61: gb\_est35:  
62: gb\_est36:  
63: gb\_est37:  
64: gb\_est38:  
65: em\_est27:  
66: em\_est28:  
67: em\_est29:  
68: em\_est30:  
69: gb\_est39:  
70: gb\_est40:  
71: gb\_est41:  
72: gb\_est42:  
73: gb\_est43:  
74: gb\_est44:  
75: em\_est31:  
76: em\_est32:  
77: em\_est33:  
78: em\_est34:  
79: gb\_est45:  
80: gb\_est46:  
81: gb\_est47:  
82: em\_est35:  
83: em\_est36:  
84: em\_est37:  
85: gb\_est48:  
86: gb\_est49:  
87: gb\_est50:  
88: gb\_est51:  
89: gb\_est52:  
90: gb\_est53:  
91: gb\_est54:  
92: gb\_est55:  
93: gb\_gss1:  
94: gb\_gss2:  
95: gb\_gss3:  
96: gb\_gss4:  
97: em\_gss1:  
98: em\_gss2:  
99: em\_gss3:  
100: em\_gss4:  
101: gb\_gss5:  
102: gb\_gss6:  
103: gb\_gss7:  
104: gb\_gss8:  
105: gb\_gss9:  
106: em\_gss5:  
107: em\_gss6:  
108: em\_gss7:  
109: em\_gss8:  
110: em\_gss9:  
111: em\_gss10:  
112: em\_gss11:  
113: gb\_gss10:  
114: gb\_gss11:  
115: em\_gss12:  
116: gb\_gss12:



```
|||||
Db 434 ATTCAACCCGCCCCCAGGTCTGTTGGT-CCCCCACCACACAGCCCTCACAGCATCC 376
QY 2090 CCATTGCTCTGCTCTTCCACCACCCCTAGTGGGACAGTGAAGGGGAGCAGTTTACCA 2149
Db 375 CCATTGCTCTGCTCTTCCACCACCCCTAGTGGGACAGTGAAGGGGAGCAGTTTACCA 316
QY 2150 GAAGATTGCTGCTGCTCTTAGGCTCTAGTCCCTCCTCAGGAATCCCTTTAGGAAGGA 2209
Db 315 GAAGATTGCTGCTGCTCTTAGGCTCTAGTCCCTCCTCAGGAATCCCTTTAGGAAGGA 256
QY 2210 CCTCAGGACACCTCTCTGCACCCCTGCTCTCTAGAGTAGTACTGCTGAGGCCCA 2269
Db 255 CCTCAGGACACCTCTCTGCACCCCTGCTCTCTAGAGTAGTACTGCTGAGGCCCA 196
QY 2270 AGTGGGTCAAAAGCAGGTATGCGCTCAGAGATGAGCTGCTGCTGGCTTTTCAGTCA 2329
Db 195 AGTGGGTCAAAAGCAGGTATGCGCTCAGAGATGAGCTGCTGCTGGCTTTTCAGTCA 136
QY 2330 GAGGGTTGGGGGTGGCCAGCCAAAGTGCCTTGCCTGGCCGCTCTTACTCCCTCCTCT 2389
Db 135 GAGGGTTGGGGGTGGCCAGCCAAAGTGCCTTGCCTGGCCGCTCTTACTCCCTCCTCT 76
QY 2390 GCTGTCTCACTTCAAGTCCATGATTTTCACTTTTCTTAATAAAGNATCAG 2441
Db 75 GCTGTCTCACTTCAAGTCCATGATTTTCACTTTTCTTAATAAAGNATCAG 24

RESULT 2
AW371381 500 bp mRNA EST 04-FEB-2000
LOCUS RC0-BT0291-011199-011-e03 BT0291 Homo sapiens cDNA, mRNA sequence.
ACCESSION AW371381
VERSION AW371381.1 GI:6876035
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE HCGP http://www.ludwig.org.br/ORESTES.
JOURNAL The FAPESP/LICR Human Cancer Genome Project
COMMENT Unpublished (1999)
On Apr 30, 1999 this sequence version replaced gi:4728180.
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?ti=RC0&t2=RC0-BT0291-
011199-011-e03&t3=1999-11-01&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 10
High quality sequence stop: 70.
Location/Qualifiers
1. .500
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="BT0291"
/dev_stage="Adult"
/note="Organ: breast; Vector: puc18; Site:1; Smal; Site:2;
SmaI; A mini-library was made by cloning products derived
from ORESTES PCR (U.S. Letters Patent application No.
196,716 - Ludwig Institute for Cancer Research) profiles
into the puc 18 vector. Reverse transcription of tissue
mRNA and cDNA amplification were performed under low
stringency conditions."

FEATURES
source
139 a 123 c 153 g 85 t
```

## ORIGIN

```
Query Match 19.6%; Score 480; DB 71; Length 500;
Best Local Similarity 98.8%; Pred. No. 6.4e-108;
Matches 494; Conservative 0; Mismatches 5; Indels 1; Gaps 1;

QY 1134 GGAGATTCGAGAAAAGTCTCTGGCACCAATCATTCAGATGTGGCAAAACAGCTGAACAA 1193
Db 1 GGAGATTCGAGAAAAGTCTCTGGCATGAATCATTCAGATGTGGCAAAACAGCTGAACAA 60
QY 1194 CTGTGGCCCTCTTGTGCCAAACAGGCGCAAGTATAGGCGCTGGAACGCTACTACAGCG 1253
Db 61 CTGTGGCCCTCTTGTGCCAAACAGGCGCAAGTATAGGCGCTGGAACGCTACTACAGCG 120
QY 1254 AGCACTGGCCATCTACGAGGGGAGCTGGGCGCGGACACCC-TAATGTAGCCGGACCA 1312
Db 121 AGCACTGGCCATCTACGAGGGGAGCTGGGCGCGGACACCC-TAATGTAGCCGGACCA 180
QY 1313 AGAACACCTGGCTTCTCTTACCTGAAACAGGGCAAAATATGCTGAGGCTGAGACACTAT 1372
Db 181 AGAACACCTGGCTTCTCTTACCTGAAACAGGGCAAAATATGCTGAGGCTGAGACACTAT 240
QY 1373 ACAAGAGATCTGACCCGCTGCCATGTACAGGAGTTTGGGTCTGTGGATGATGACACCA 1432
Db 241 ACAGAGAGATCTGACCCGCTGCCATGTACAGGAGTTTGGGTCTGTGGATGATGACACCA 300
QY 1433 AGCCCATCTGGATGCATGCAGAGGAGCGGAGAAATGAGCAAAAGCCGCACCATGAGG 1492
Db 301 AGCCCATCTGGATGCATGCAGAGGAGCGGAGAAATGAGCAAAAGCCGCACCATGAGG 360
QY 1493 GTGGGACACCTATGCTGAGTATGGAGGCTGTACAGGCTCTCAAAGTGGAGAGGCCCA 1552
Db 361 GTGGGACACCTATGCTGAGTATGGAGGCTGTACAGGCTCTCAAAGTGGAGAGGCCCA 420
QY 1553 CAGTGAACACTACTCTGAGAAACCTGGAGCTCTGTATAGGCCCGAGGAAAGCTGGAGG 1612
Db 421 CAGTGAACACTACTCTGAGAAACCTGGAGCTCTGTATAGGCCCGAGGAAAGCTGGAGG 480
QY 1613 CTGCTGAGACCCCTGGAGGA 1632
Db 481 CTGCTGAGACCCCTGGAGGA 500
```

## RESULT 3

## AI041608/c

## LOCUS

## DEFINITION

## AI041608

## ACCESSION

## AI041608

## VERSION

## AI041608.1

## GI:3280802

## KEYWORDS

## EST.

## SOURCE

## human.

## ORGANISM

## Homo sapiens

## Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

## Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

## REFERENCE

## 1 (bases 1 to 463)

## AUTHORS

## NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

## TITLE

## National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index

## Unpublished (1997)

## On Jun 5, 1998 this sequence version replaced gi:3187022.

## Contact: Robert Strausberg, Ph.D.

## Tel: (301) 496-1550

## Email: Robert.Strausberg@nih.gov

## This clone is available royalty-free through LLNL; contact the

## IMAGE Consortium (info@image.llnl.gov) for further information.

## Insert Length: 972 Std Error: 0.00

## Seq primer: -40ml3 fwd. ET from Amersham

## High quality sequence stop: 419.

## Location/Qualifiers

## 1. .463

## /organism="Homo sapiens"

## source

## 139 a 123 c 153 g 85 t

```
/db_xref="taxon:9606"
/clone_image="1663614"
/clone_lib="Soares_senescent_fibroblasts_NDHSF"
/tissue_type="senescent_fibroblast"
/lab_host="DH10B (ampicillin resistant)"
/notes="Vector: pT73D (Pharmacia) with a modified
polylinker V-TYPE: phagemid; Site_1: Not I; Site_2: Eco
RI; 1st strand cDNA was primed with a Not I - oligo(dT)
primer [5']
TGTACCAATCTGAAGTGGAGCGCGCATTTTTTTTTTTTTTTT 3']
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT73 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by Bento
Soares and M.Fatima Bonaldo."
BASE COUNT      106 a 107 c 154 g
ORIGIN

Query Match      18.4%; Score 451; DB 34; Length 463;
Best Local Similarity 100.0%; Pred. No. 8.8e-101;
Matches 451; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1991 TCAGTGCAGCAGCATGGAGCTCTCTTCAAGCAGCTGACATTCACCCGGCCCCCAGGTC 2050
DB 463 TCAGTGCAGCAGCATGGAGCTCTCTTCAAGCAGCTGACATTCACCCGGCCCCCAGGTC 404

QY 2051 TGCTGGGTGCCCCCACCACCCACAGCCCTCAGCATTCCTCCCTGCTGGCTCTTCC 2110
DB 403 TGCTGGGTGCCCCCACCACCCACAGCCCTCAGCATTCCTCCCTGCTGGCTCTTCC 344

QY 2111 CACCCCTAGTGGGACAGTGAAGGGGAGCAGTTTAAACAGAGATTGCTGCTCCCTTAG 2170
DB 343 CACCCCTAGTGGGACAGTGAAGGGGAGCAGTTTAAACAGAGATTGCTGCTCCCTTAG 284

QY 2171 GGCTCAGCTCCCTCCTCAGGAATCCCTCTTAGAAGGACCCCTCAGGACACCTCTCTGC 2230
DB 283 GGCTCAGCTCCCTCCTCAGGAATCCCTCTTAGAAGGACCCCTCAGGACACCTCTCTGC 224

QY 2231 ACCCTGTGGTCTCTAGTAGTGTAGTCTGAGGCCCAAGTGGGTACAAAGCAGGTAT 2290
DB 223 ACCCTGTGGTCTCTAGTAGTGTAGTCTGAGGCCCAAGTGGGTACAAAGCAGGTAT 164

QY 2291 GGCCCTCAGAGAGCAGCCCTGCTGCTGGCTTTTCAGTCAGAGGGTTGGGGGCTGGCCAGC 2350
DB 163 GGCCCTCAGAGAGCAGCCCTGCTGCTGGCTTTTCAGTCAGAGGGTTGGGGGCTGGCCAGC 104

QY 2351 CAAGCTGCCTTGGCTGGCGCTTACTCCCTCCCTCTGCTGCTCAGTTCAGTTCAT 2410
DB 103 CAAGCTGCCTTGGCTGGCGCTTACTCCCTCCCTCTGCTGCTCAGTTCAGTTCAT 44

QY 2411 GTATTTCACTTTTCTTAATAAAGAATCAG 2441
DB 43 GTATTTCACTTTTCTTAATAAAGAATCAG 13

RESULT 4
AI814907/c
LOCUS      747 bp      mRNA      EST      07-MAR-2000
DEFINITION      w68e02.x1 NCI_CGAP_Pan1 Homo sapiens cDNA clone IMAGE:2420570 3',
                mRNA sequence.
ACCESSION      AI814907
VERSION        AI814907.1 GI:5426122
KEYWORDS       EST.
SOURCE         human.
ORGANISM       Homo sapiens
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
                1 (bases 1 to 747)
REFERENCE      NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
AUTHORS       National Cancer Institute, Cancer Genome Anatomy Project (CGAP).
TITLE         Tumor Gene Index
```

```
Unpublished (1997)
On Jun 15, 1998 this sequence version replaced gi:3226427.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Life Technologies catalog #: 11548-013
DNA sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 899 Std Error: 0.00
Seq Primer: -400P from Gibco
High quality sequence stop: 390.
Location/Qualifiers
1..747
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="2420570"
/clone_lib="NCI_CGAP_Pan1"
/tissue_type="adenocarcinoma"
/lab_host="DH10B"
/notes="Organ: pancreas; Vector: pCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primers: Oligo dT.
Average insert size 1.72 Kb. Life technologies catalog #:
11548-013"
BASE COUNT      159 a 164 c 233 g 167 t 24 others
ORIGIN

Query Match      17.7%; Score 434.4; DB 44; Length 747;
Best Local Similarity 93.7%; Pred. No. 1.2e-96;
Matches 503; Conservative 0; Mismatches 28; Indels 6; Gaps 5;

QY 1905 GCGCTCCACAGCAACATGAAGCGAGCAGCTCTTGAACCTATCTGAACCAACCTAGTGC 1964
DB 540 GCGCTCAGAGACACAT-AAGCGAGCAGCTTCTTGAACCTATCTGAAC--AACTTAGTC 484

QY 1965 AGCACCCCTCAGGTCTCCCGGGGCTCAGTGCAGCAGCACCACCTCTCTTCAAGCAG 2024
DB 483 AGCACCCCTCAGGTCTCCA-GGGCTAAGTCAAGCACAAT-GACCTCTTCAAGCAG 426

QY 2025 CTGACATTCACCCGGCCCCCAGGTCTGCTGGGTGTCGCCACCCACAGCCCTCAGCAG 2084
DB 425 NTACAAATTCACCGGCCNNCAGGTGNTGGNT-CCCCCACCACCCACAGCCCTCAGCAG 367

QY 2085 ATTCCCAATTCCTGCTGCTCTTCCACCCCTAGTGGGACAGTGAAGGGAGCAGTTT 2144
DB 366 ATTCCCAATTCCTGCTGCTCTTCCACCCCTAGTGGGACAGTGAAGGGAGCAGTTT 307

QY 2145 AACCAAGAATTGCTGCTGCTTGGGTCTCAGCTCCCTCCTCAGGAATCCCTCTTAGG 2204
DB 306 AACCAAGAATTGCTGCTGCTTGGGTCTCAGCTCCCTCCTCAGGAATCCCTCTTAGG 247

QY 2205 AAGGACCTCAGGACACCTCTCTGACCCCTGCTGCTCTAGAGTAGCTAGCTTAGG 2264
DB 246 AAGGACCTCAGGACACCTCTCTGACCCCTGCTGCTCTAGAGTAGCTAGCTTAGG 187

QY 2265 CCCCAGGTGGGTACAAAGCAGGTATGGCCCTCAGAGATGAGCTGCTGCTGCTTTTC 2324
DB 186 CCCCAGGTGGGTACAAAGCAGGTATGGCCCTCAGAGATGAGCTGCTGCTGCTTTTC 127

QY 2325 AGTCAGAGGTTGGGGGCTGGCCAGCAAGCTGCTTGGCCCTGCGCTCTTACTCCCTC 2384
DB 126 AGTCAGAGGTTGGGGGCTGGCCAGCAAGCTGCTTGGCCCTGCGCTCTTACTCCCTC 67

QY 2385 CCTGCTGCTCAGTTCAGGTCCATGATTTTCACTTTTCTTAATAAAGAATCAG 2441
DB 66 CCTGCTGCTCAGTTCAGGTCCATGATTTTCACTTTTCTTAATAAAGAATCAG 10

RESULT 5
AI076952/c
LOCUS      450 bp      mRNA      EST      27-AUG-1998
```

DEFINITION ov20h09.xl NCI\_CGAP\_Br2 Homo sapiens cDNA clone IMAGE:1637921 3', mRNA sequence.  
ACCESSION AI076952  
VERSION AI076952.1 GI:3404781  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
AUTHORS 1 (bases 1 to 450)  
TITLE NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
JOURNAL National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index  
COMMENT Unpublished (1997)  
On Jan 19, 1998 this sequence version replaced gi:2285573.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert\_Strausberg@nih.gov  
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.  
CDNA Library Preparation: M. Bento Soares, Ph.D.  
CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html  
Insert Length: 614 Std Error: 0.00  
Seq primer: -40ml3 fwd. ET from Amersham.  
FEATURES  
Source  
1..450  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1637921"  
/clone.lib="NCI\_CGAP\_Br2"  
/sex="female, pooled"  
/tissue\_type="breast"  
/lab\_host="DH10B"  
/note="Vector: p7T3D-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from pooled bulk breast tumor tissue, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified p7T3 vector. This library is the normalized version of NCI\_CGAP\_Br1.1. Library was constructed by Bento Soares and M. Fatima Bonaldo."  
BASE COUNT 104 a 102 c 147 g 96 t 1 others  
ORIGIN  
Query Match 17.4%; Score 428; DB 35; Length 450;  
Best Local Similarity 99.8%; Pred. No. 4, 1e-95;  
Matches 428; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 2013 CTCCTTCAAGCAGCTGACATTCACCCGCCGCCAGTCTGCTGGTCCGCCACCCAC 2072  
DB 450 CTCCTTCAAGCAGCTGACATTCACCCGCCGCCAGTCTGCTGGTCCGCCACCCAC 391  
QY 2073 AGCCCTCACAGATTCCTCCATTCCTGGCTCTTCCGCCACCCCTAGGTGGACAGTGAA 2132  
DB 390 AGCCCTCACAGATTCCTCCATTCCTGGCTCTTCCGCCACCCCTAGGTGGACAGTGAA 331  
QY 2133 GGGGAGCAGTTTAAACAGAGATTCCTGCTGCCCTTAGGTCTCAGCTCCCTCCACAGA 2192  
DB 330 GGGGAGCAGTTTAAACAGAGATTCCTGCTGCCCTTAGGTCTCAGCTCCCTCCACAGA 271  
QY 2193 ATCCCTCTTAGGAGACCCCTCAGGACACCCCTCTCTGCACCCCTGGTCTCTAGAGTAG 2252  
DB 270 ATCCCTCTTAGGAGACCCCTCAGGACACCCCTCTCTGCACCCCTGGTCTCTAGAGTAG 211  
QY 2253 CTAGCTCTGAGGCCCCCAGGTGGGTACAAAGCAGGTATGGCCCTCAGAGATCAGCCTGC 2312  
DB 210 CTAGCTCTGAGGCCCCCAGGTGGGTACAAAGCAGGTATGGCCCTCAGAGATCAGCCTGC 151

QY 2313 TGCTGGCTTTTCAGTCAGAGGTTGGGGCTGGCCAGCAAGCTGCCTTGCCTGGCCGC 2372  
DB 150 TGCTGGCTTTTCAGTCAGAGGTTGGGGCTGGCCAGCAAGCTGCCTTGCCTGGCCGC 91  
QY 2373 TCTTACTCCCTCCCTCTGCTCTCCTCCTCCTCCTCCTCCTCCTCCTCCTTCTTAATAA 2432  
DB 90 TCTTACTCCCTCCCTCTGCTCTCCTCCTCCTCCTCCTCCTCCTCCTTCTTAATAA 31  
QY 2433 AAGAATCAG 2441  
DB 30 AAGAATCAG 22  
RESULT 6  
AL184140/C  
LOCUS AL184140 441 bp mRNA EST 10-NOV-1998  
DEFINITION qf30b02.xl Soares\_testis\_NHT Homo sapiens cDNA clone IMAGE:1751499  
ACCESSION AL184140  
VERSION AL184140.1 GI:3734778  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
AUTHORS 1 (bases 1 to 441)  
TITLE NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
JOURNAL National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index  
COMMENT Unpublished (1997)  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert\_Strausberg@nih.gov  
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.  
CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html  
Insert Length: 1426 Std Error: 0.00  
Seq primer: -40UP from Gibco  
High quality sequence stop: 431.  
FEATURES  
Source  
1..441  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1751499"  
/clone.lib="Soares\_testis\_NHT"  
/sex="male"  
/lab\_host="DH10B"  
/note="Vector: p7T3D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was prepared from mRNA obtained from Clontech Laboratories, Inc., and primed with a Not I - oligo(dT) primer [5', TGTTACCAATCTGAAGTGGGAGCGCGCCCAATTTTTTTTTTTT 3']. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified p7T3 vector. Library went through one round of normalization to Cot5, and was constructed by Bento Soares and M. Fatima Bonaldo."  
BASE COUNT 105 a 102 c 148 g 85 t 1 others  
ORIGIN  
Query Match 17.4%; Score 427.4; DB 36; Length 441;  
Best Local Similarity 99.5%; Pred. No. 5, 7e-95;  
Matches 428; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 2012 TCTTTCAGCAGCTGACATTCACCCGCCGCCAGTCTGCTGGTCCGCCACCCCA 2071

```

Db 441 TCCTTCAACGAGCTGACATTCACCGGGCCCGCCAGGTCTGNTGGTCCCGCCACCCCA 382
QY 2072 CAGCCCTCACAGATTCCTCCCTGCTCTTCCACCCCTAGGTGGGACAGTGA 2131
Db 381 CAGCCCTCACAGATTCCTCCCTGCTCTTCCACCCCTAGGTGGGACAGTGA 322
QY 2132 AGGGAGCAGTTTAAACAGAGATTCCTGCTGCCCTTAGGTCTCAGTCCCTCCTCAGG 2191
Db 321 AGGGAGCAGTTTAAACAGAGATTCCTGCTGCCCTTAGGTCTCAGTCCCTCCTCAGG 262
QY 2192 AATCCCTCTTAGGAAGACCTCAGACACCCCTCTCTGACCCCTGCTGCTCTTAGAGTA 2251
Db 261 AATCCCTCTTAGGAAGACCTCAGACACCCCTCTCTGACCCCTGCTGCTCTTAGAGTA 202
QY 2252 GCTAGCTCTGAGCCCAAGGTGGGTACAAAGCAGGTATGGCCCTCAGAGATGCAGCCTG 2311
Db 201 GCTAGCTCTGAGCCCAAGGTGGGTACAAAGCAGGTATGGCCCTCAGAGATGCAGCCTG 142
QY 2312 CTGCTGGCTTTTTCAGTCAGAGGTTGGGGCTGGCCAGCAGCTGCCCTGGCCGG 2371
Db 141 CTGCTGGCTTTTTCAGTCAGAGGTTGGGGCTGGCCAGCAGCTGCCCTGGCCGG 82
QY 2372 CTCTTACTCCCTCCCTGCTGCTCCTCCTCAGTCCATGTATTTTCAATA 2431
Db 81 CTCTTACTCCCTCCCTGCTGCTCCTCCTCAGTCCATGTATTTTCAATA 22
QY 2432 AAGAATCAG 2441
Db 21 AAGAATCAG 12

RESULT 7
AW072690/c
LOCUS
DEFINITION
  xa41h04.x1 NCI_CGAP_Sar4 Homo sapiens cDNA clone IMAGE:2569399 3',
  mRNA sequence.
ACCESSION
  AW072690
VERSION
  AW072690.1 GI:6027688
KEYWORDS
  EST.
SOURCE
  human.
ORGANISM
  Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
  1 (bases 1 to 758)
  NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
  Tumor Gene Index
  Unpublished (1997)
  On Jun 5, 1998 this sequence version replaced gi:3189325.
  Contact: Robert Strausberg, Ph.D.
  Tel: (301) 496-1550
  Email: Robert.Strausberg@nih.gov
  Tissue Procurement: Chris Moskaluk, M.D., Ph.D., Michael R.
  Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: Life
  Technologies, Inc. cDNA Library Arrayed by: Christa Frange, The
  I.M.A.G.E. Consortium DNA Sequencing by: Washington University
  Genome Sequencing Center
  Clone distribution: NCI-CGAP clone distribution information can be
  found through the I.M.A.G.E. Consortium/LLNL at:
  www-bio.llnl.gov/bbrp/image/image.html

Possible reversed clone: polyT not found
Seq primer: -40UP from Gibco
High quality sequence stop: 389.
  Location/Qualifiers
    1..758
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /clone="IMAGE:2569399"
      /clone_lib="NCI_CGAP_Sar4"
      /tissue_type="five pooled sarcomas, including myxoid
      liposarcoma, solitary fibrous tumor, malignant fibrous
      histiocytoma, gastrointestinal stromal tumor, and

```

```

mesothelioma"
/lab_host="DH10B"
/notes="Organ: connective tissue; Vector: pCMV-SPORT6;
Site1: SalI; Site2: NotI; Cloned unidirectionally.
Primer: Oligo dt. Library constructed by Life
Technologies."
BASE COUNT 180 a 157 c 261 g 152 t 8 others
ORIGIN

Query Match
Best Local Similarity 17.4%; Score 426.6; DB 63; Length 758;
Matches 450; Conservative 0; Mismatches 25; Indels 2; Gaps 1;

QY 1965 AGCACCCCTCCAGTCTCCCGGGGCTCAGTGCACGACCATGACCTCTCTTCAAGCAG 2024
Db 485 AGCCCCCTTCAGGTCTTC--GTGGCTTCAGTGCAGGCCCATGACCTCTCTCNAGCAG 428
QY 2025 CTGACATTCACCGGGCCCGCCAGGTCTGCTGGGTGCCCGCCACCCACACGCCCTCACAGC 2084
Db 427 CTGACATTCACCGGGCCCGCCANNNGTNTGTTGGTCCCNCCACCCACAGCCCTCACAGC 368
QY 2085 ATTCGCCATTGCTCTCGGCTCTTCCCGACCCCTAGGTGGGACAGTGAAGGGGAGCAGTTT 2144
Db 367 ATTCGCCATTGCTCTCGGCTCTTCCCGACCCCTAGGTGGGACAGTGAAGGGGAGCAGTTT 308
QY 2145 AACCAGAAGATTGCTGCTGCCCTTAGGGTCTCAGCTCCCTCCTCAGGAATCCCTCTTAGG 2204
Db 307 AACCAGAAGATTGCTGCTGCCCTTAGGGTCTCAGCTCCCTCCTCAGGAATCCCTCTTAGG 248
QY 2205 AAGACCTCAGGACACCCCTCTCTGCACCTCTGGTCTCTAGAGTAGTAGTCTGAGG 2264
Db 247 AAGACCTCAGGACACCCCTCTCTGCACCTCTGGTCTCTAGAGTAGTAGTCTGAGG 188
QY 2265 CCCCAAGGTGGGTACAAAGCAGGTATGGCCCTCAGAGATGCAGCTGCTGGCTTTTC 2324
Db 187 CCCCAAGGTGGGTACAAAGCAGGTATGGCCCTCAGAGATGCAGCTGCTGGCTTTTC 128
QY 2325 AGTCAGAGGTTGGGGGCTGGCCAGCAAGCTGCCCTGGCCGCTCTTACTCCCTC 2384
Db 127 AGTCAGAGGTTGGGGGCTGGCCAGCAAGCTGCCCTGGCCGCTCTTACTCCCTC 68
QY 2385 CCTCTGCTGCTCACTTCAGGTCCATGTATTTTCAATTAAGAATCAG 2441
Db 67 CCTCTGCTGCTCACTTCAGGTCCATGTATTTTCAATTAAGAATCAG 11

RESULT 8
AA678056/c
LOCUS
DEFINITION
  2125f04.s1 Soares_fetal_liver_spleen_infls_s1 Homo sapiens cDNA
  clone IMAGE:431839 3', mRNA sequence.
ACCESSION
  AA678056
VERSION
  AA678056.1 GI:2658578
KEYWORDS
  EST.
SOURCE
  human.
ORGANISM
  Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
  1 (bases 1 to 425)
  Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,
  Krizman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M.,
  Martin, J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F.,
  Theising, B., White, Y., Wyllie, T., Waterston, R. and Wilson, R.
  WashU-NCI human EST Project
  Unpublished (1997)
  Contact: Wilson RK
  Washington University School of Medicine
  4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
  Tel: 314 286 1800
  Fax: 314 286 1810
  Email: est@watson.wustl.edu
  This clone is available royalty-free through LxNL; contact the

```



IMAGE Consortium (info@image.llnl.gov) for further information.	
Seq primer: -40m13 fwd. ER from Amersham	
High quality sequence stop: 410.	
Location/Qualifiers	
1. 425	
/organism="Homo sapiens"	
/db_xref="taxon:9606"	
/clone="IMAGE:431839"	
/clone_lib="Soares fetal_liver_spleen_INFLS_S1"	
/sex="male"	
/dev_stage="20 week-post conception fetus"	
/lab_host="DH10B (ampicillin resistant)"	
/notes="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia) with a modified polylinker; Site_1: Pac I; Site_2: Eco RI; This is a subcloned version of the original Soares fetal liver spleen INFLS library. 1st strand cDNA was primed with a Pac I - oligo(dT) primer [5', AACTGGAAGTAATTAATGATCTTTTTTTTTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac I and cloned into the Pac I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fatima Bonaldo."	
99 a 101 c 143 g 82 t	
BASE COUNT	

BASE COUNT	99 a	101 c	143 g	82 t	Constructed by Benito Soares and M.Fátima Bonardão.																																																																																																																																																																		
ORIGIN																																																																																																																																																																							
Query Match	16.78;	Score	410;	DB 29;	Length 425;																																																																																																																																																																		
Best Local Similarity	99.8;	Pred. No.	1.1e-90;																																																																																																																																																																				
Matches 421; Conservative	0;	Mismatches	0;	Indels	1; Gaps																																																																																																																																																																		
1; 1;																																																																																																																																																																							
QY	2020	AGCAGCTGACATTCAACCCCGGCCCCAGGTCGTGCTGGTCCCCACACCCACAGCCCTC	2079																																																																																																																																																																				
						QY	2080	ACAGCATTCOCATTCGCTCGGCTCTTCGCCACCCCTAGGTGGGACAGTGAAGGGGAGC	2139									Db	366	ACAGCATTCOCATTCGCTCGGCTCTTCGCCACCCCTAGGTGGGACAGTGAAGGGGAGC	307									QY	2140	AGTTTACCAGAAATGCTGCTGCCCTTAGGGTCTCAGCTCCCTCCTCAGGAATCCCTC	2199									Db	306	AGTTTACCAGAAATGCTGCTGCCCTTAGGGTCTCAGCTCCCTCCTCAGGAATCCCTC	247									QY	2200	TTAGGAAGGACCCCTCAGGACACCCCTCTCTGCACCCCTGFGTCCCTCTAGAGTAGCTAGCTC	2259									Db	246	TTAGGAAGGACCCCTCAGGACACCCCTCTCTGCACCCCTGFGTCCCTCTAGAGTAGCTAGCTC	187									QY	2260	TGAGGCCCAAGGTGGGTACAAAGCAGGTATGGCCCTCAGAGATGCAGCCCTGCTGCTGGC	2319									Db	186	TGAGGCCCAAGGTGGGTACAAAGCAGGTATGGCCCTCAGAGATGCAGCCCTGCTGCTGGC	127									QY	2320	TTTTTCAGTCAGAGGTTGGGGGCTGGCCAGCCAGCTGCCTTTGCCCTGGCCGCTCTTACT	2379									Db	126	TTTTTCAGTCAGAGGTTGGGGGCTGGCCAGCCAGCTGCCTTTGCCCTGGCCGCTCTTACT	67									QY	2380	CCCTCCCTCTGCTCTCTCAGCTTCAGGTCCTATTTTTCACCTTTTCTTAAATAAAGAAATC	2439									Db	66	CCCTCCCTCTGCTCTCTCAGCTTCAGGTCCTATTTTTCACCTTTTCTTAAATAAAGAAATC	7									QY	2440	AG 2441										Db	6	AG 5			
QY	2080	ACAGCATTCOCATTCGCTCGGCTCTTCGCCACCCCTAGGTGGGACAGTGAAGGGGAGC	2139									Db	366	ACAGCATTCOCATTCGCTCGGCTCTTCGCCACCCCTAGGTGGGACAGTGAAGGGGAGC	307									QY	2140	AGTTTACCAGAAATGCTGCTGCCCTTAGGGTCTCAGCTCCCTCCTCAGGAATCCCTC	2199									Db	306	AGTTTACCAGAAATGCTGCTGCCCTTAGGGTCTCAGCTCCCTCCTCAGGAATCCCTC	247									QY	2200	TTAGGAAGGACCCCTCAGGACACCCCTCTCTGCACCCCTGFGTCCCTCTAGAGTAGCTAGCTC	2259									Db	246	TTAGGAAGGACCCCTCAGGACACCCCTCTCTGCACCCCTGFGTCCCTCTAGAGTAGCTAGCTC	187									QY	2260	TGAGGCCCAAGGTGGGTACAAAGCAGGTATGGCCCTCAGAGATGCAGCCCTGCTGCTGGC	2319									Db	186	TGAGGCCCAAGGTGGGTACAAAGCAGGTATGGCCCTCAGAGATGCAGCCCTGCTGCTGGC	127									QY	2320	TTTTTCAGTCAGAGGTTGGGGGCTGGCCAGCCAGCTGCCTTTGCCCTGGCCGCTCTTACT	2379									Db	126	TTTTTCAGTCAGAGGTTGGGGGCTGGCCAGCCAGCTGCCTTTGCCCTGGCCGCTCTTACT	67									QY	2380	CCCTCCCTCTGCTCTCTCAGCTTCAGGTCCTATTTTTCACCTTTTCTTAAATAAAGAAATC	2439									Db	66	CCCTCCCTCTGCTCTCTCAGCTTCAGGTCCTATTTTTCACCTTTTCTTAAATAAAGAAATC	7									QY	2440	AG 2441										Db	6	AG 5									
						Db	366	ACAGCATTCOCATTCGCTCGGCTCTTCGCCACCCCTAGGTGGGACAGTGAAGGGGAGC	307									QY	2140	AGTTTACCAGAAATGCTGCTGCCCTTAGGGTCTCAGCTCCCTCCTCAGGAATCCCTC	2199									Db	306	AGTTTACCAGAAATGCTGCTGCCCTTAGGGTCTCAGCTCCCTCCTCAGGAATCCCTC	247									QY	2200	TTAGGAAGGACCCCTCAGGACACCCCTCTCTGCACCCCTGFGTCCCTCTAGAGTAGCTAGCTC	2259									Db	246	TTAGGAAGGACCCCTCAGGACACCCCTCTCTGCACCCCTGFGTCCCTCTAGAGTAGCTAGCTC	187									QY	2260	TGAGGCCCAAGGTGGGTACAAAGCAGGTATGGCCCTCAGAGATGCAGCCCTGCTGCTGGC	2319									Db	186	TGAGGCCCAAGGTGGGTACAAAGCAGGTATGGCCCTCAGAGATGCAGCCCTGCTGCTGGC	127									QY	2320	TTTTTCAGTCAGAGGTTGGGGGCTGGCCAGCCAGCTGCCTTTGCCCTGGCCGCTCTTACT	2379									Db	126	TTTTTCAGTCAGAGGTTGGGGGCTGGCCAGCCAGCTGCCTTTGCCCTGGCCGCTCTTACT	67									QY	2380	CCCTCCCTCTGCTCTCTCAGCTTCAGGTCCTATTTTTCACCTTTTCTTAAATAAAGAAATC	2439									Db	66	CCCTCCCTCTGCTCTCTCAGCTTCAGGTCCTATTTTTCACCTTTTCTTAAATAAAGAAATC	7									QY	2440	AG 2441										Db	6	AG 5															
Db	366	ACAGCATTCOCATTCGCTCGGCTCTTCGCCACCCCTAGGTGGGACAGTGAAGGGGAGC	307									QY	2140	AGTTTACCAGAAATGCTGCTGCCCTTAGGGTCTCAGCTCCCTCCTCAGGAATCCCTC	2199									Db	306	AGTTTACCAGAAATGCTGCTGCCCTTAGGGTCTCAGCTCCCTCCTCAGGAATCCCTC	247									QY	2200	TTAGGAAGGACCCCTCAGGACACCCCTCTCTGCACCCCTGFGTCCCTCTAGAGTAGCTAGCTC	2259									Db	246	TTAGGAAGGACCCCTCAGGACACCCCTCTCTGCACCCCTGFGTCCCTCTAGAGTAGCTAGCTC	187									QY	2260	TGAGGCCCAAGGTGGGTACAAAGCAGGTATGGCCCTCAGAGATGCAGCCCTGCTGCTGGC	2319									Db	186	TGAGGCCCAAGGTGGGTACAAAGCAGGTATGGCCCTCAGAGATGCAGCCCTGCTGCTGGC	127									QY	2320	TTTTTCAGTCAGAGGTTGGGGGCTGGCCAGCCAGCTGCCTTTGCCCTGGCCGCTCTTACT	2379									Db	126	TTTTTCAGTCAGAGGTTGGGGGCTGGCCAGCCAGCTGCCTTTGCCCTGGCCGCTCTTACT	67									QY	2380	CCCTCCCTCTGCTCTCTCAGCTTCAGGTCCTATTTTTCACCTTTTCTTAAATAAAGAAATC	2439									Db	66	CCCTCCCTCTGCTCTCTCAGCTTCAGGTCCTATTTTTCACCTTTTCTTAAATAAAGAAATC	7									QY	2440	AG 2441										Db	6	AG 5																					
						QY	2140	AGTTTACCAGAAATGCTGCTGCCCTTAGGGTCTCAGCTCCCTCCTCAGGAATCCCTC	2199									Db	306	AGTTTACCAGAAATGCTGCTGCCCTTAGGGTCTCAGCTCCCTCCTCAGGAATCCCTC	247									QY	2200	TTAGGAAGGACCCCTCAGGACACCCCTCTCTGCACCCCTGFGTCCCTCTAGAGTAGCTAGCTC	2259									Db	246	TTAGGAAGGACCCCTCAGGACACCCCTCTCTGCACCCCTGFGTCCCTCTAGAGTAGCTAGCTC	187									QY	2260	TGAGGCCCAAGGTGGGTACAAAGCAGGTATGGCCCTCAGAGATGCAGCCCTGCTGCTGGC	2319									Db	186	TGAGGCCCAAGGTGGGTACAAAGCAGGTATGGCCCTCAGAGATGCAGCCCTGCTGCTGGC	127									QY	2320	TTTTTCAGTCAGAGGTTGGGGGCTGGCCAGCCAGCTGCCTTTGCCCTGGCCGCTCTTACT	2379									Db	126	TTTTTCAGTCAGAGGTTGGGGGCTGGCCAGCCAGCTGCCTTTGCCCTGGCCGCTCTTACT	67									QY	2380	CCCTCCCTCTGCTCTCTCAGCTTCAGGTCCTATTTTTCACCTTTTCTTAAATAAAGAAATC	2439									Db	66	CCCTCCCTCTGCTCTCTCAGCTTCAGGTCCTATTTTTCACCTTTTCTTAAATAAAGAAATC	7									QY	2440	AG 2441										Db	6	AG 5																											
QY	2140	AGTTTACCAGAAATGCTGCTGCCCTTAGGGTCTCAGCTCCCTCCTCAGGAATCCCTC	2199									Db	306	AGTTTACCAGAAATGCTGCTGCCCTTAGGGTCTCAGCTCCCTCCTCAGGAATCCCTC	247									QY	2200	TTAGGAAGGACCCCTCAGGACACCCCTCTCTGCACCCCTGFGTCCCTCTAGAGTAGCTAGCTC	2259									Db	246	TTAGGAAGGACCCCTCAGGACACCCCTCTCTGCACCCCTGFGTCCCTCTAGAGTAGCTAGCTC	187									QY	2260	TGAGGCCCAAGGTGGGTACAAAGCAGGTATGGCCCTCAGAGATGCAGCCCTGCTGCTGGC	2319									Db	186	TGAGGCCCAAGGTGGGTACAAAGCAGGTATGGCCCTCAGAGATGCAGCCCTGCTGCTGGC	127									QY	2320	TTTTTCAGTCAGAGGTTGGGGGCTGGCCAGCCAGCTGCCTTTGCCCTGGCCGCTCTTACT	2379									Db	126	TTTTTCAGTCAGAGGTTGGGGGCTGGCCAGCCAGCTGCCTTTGCCCTGGCCGCTCTTACT	67									QY	2380	CCCTCCCTCTGCTCTCTCAGCTTCAGGTCCTATTTTTCACCTTTTCTTAAATAAAGAAATC	2439									Db	66	CCCTCCCTCTGCTCTCTCAGCTTCAGGTCCTATTTTTCACCTTTTCTTAAATAAAGAAATC	7									QY	2440	AG 2441										Db	6	AG 5																																	
						Db	306	AGTTTACCAGAAATGCTGCTGCCCTTAGGGTCTCAGCTCCCTCCTCAGGAATCCCTC	247									QY	2200	TTAGGAAGGACCCCTCAGGACACCCCTCTCTGCACCCCTGFGTCCCTCTAGAGTAGCTAGCTC	2259									Db	246	TTAGGAAGGACCCCTCAGGACACCCCTCTCTGCACCCCTGFGTCCCTCTAGAGTAGCTAGCTC	187									QY	2260	TGAGGCCCAAGGTGGGTACAAAGCAGGTATGGCCCTCAGAGATGCAGCCCTGCTGCTGGC	2319									Db	186	TGAGGCCCAAGGTGGGTACAAAGCAGGTATGGCCCTCAGAGATGCAGCCCTGCTGCTGGC	127									QY	2320	TTTTTCAGTCAGAGGTTGGGGGCTGGCCAGCCAGCTGCCTTTGCCCTGGCCGCTCTTACT	2379									Db	126	TTTTTCAGTCAGAGGTTGGGGGCTGGCCAGCCAGCTGCCTTTGCCCTGGCCGCTCTTACT	67									QY	2380	CCCTCCCTCTGCTCTCTCAGCTTCAGGTCCTATTTTTCACCTTTTCTTAAATAAAGAAATC	2439									Db	66	CCCTCCCTCTGCTCTCTCAGCTTCAGGTCCTATTTTTCACCTTTTCTTAAATAAAGAAATC	7									QY	2440	AG 2441										Db	6	AG 5																																							
Db	306	AGTTTACCAGAAATGCTGCTGCCCTTAGGGTCTCAGCTCCCTCCTCAGGAATCCCTC	247									QY	2200	TTAGGAAGGACCCCTCAGGACACCCCTCTCTGCACCCCTGFGTCCCTCTAGAGTAGCTAGCTC	2259									Db	246	TTAGGAAGGACCCCTCAGGACACCCCTCTCTGCACCCCTGFGTCCCTCTAGAGTAGCTAGCTC	187									QY	2260	TGAGGCCCAAGGTGGGTACAAAGCAGGTATGGCCCTCAGAGATGCAGCCCTGCTGCTGGC	2319									Db	186	TGAGGCCCAAGGTGGGTACAAAGCAGGTATGGCCCTCAGAGATGCAGCCCTGCTGCTGGC	127									QY	2320	TTTTTCAGTCAGAGGTTGGGGGCTGGCCAGCCAGCTGCCTTTGCCCTGGCCGCTCTTACT	2379									Db	126	TTTTTCAGTCAGAGGTTGGGGGCTGGCCAGCCAGCTGCCTTTGCCCTGGCCGCTCTTACT	67									QY	2380	CCCTCCCTCTGCTCTCTCAGCTTCAGGTCCTATTTTTCACCTTTTCTTAAATAAAGAAATC	2439									Db	66	CCCTCCCTCTGCTCTCTCAGCTTCAGGTCCTATTTTTCACCTTTTCTTAAATAAAGAAATC	7									QY	2440	AG 2441										Db	6	AG 5																																													
						QY	2200	TTAGGAAGGACCCCTCAGGACACCCCTCTCTGCACCCCTGFGTCCCTCTAGAGTAGCTAGCTC	2259									Db	246	TTAGGAAGGACCCCTCAGGACACCCCTCTCTGCACCCCTGFGTCCCTCTAGAGTAGCTAGCTC	187									QY	2260	TGAGGCCCAAGGTGGGTACAAAGCAGGTATGGCCCTCAGAGATGCAGCCCTGCTGCTGGC	2319									Db	186	TGAGGCCCAAGGTGGGTACAAAGCAGGTATGGCCCTCAGAGATGCAGCCCTGCTGCTGGC	127									QY	2320	TTTTTCAGTCAGAGGTTGGGGGCTGGCCAGCCAGCTGCCTTTGCCCTGGCCGCTCTTACT	2379									Db	126	TTTTTCAGTCAGAGGTTGGGGGCTGGCCAGCCAGCTGCCTTTGCCCTGGCCGCTCTTACT	67									QY	2380	CCCTCCCTCTGCTCTCTCAGCTTCAGGTCCTATTTTTCACCTTTTCTTAAATAAAGAAATC	2439									Db	66	CCCTCCCTCTGCTCTCTCAGCTTCAGGTCCTATTTTTCACCTTTTCTTAAATAAAGAAATC	7									QY	2440	AG 2441										Db	6	AG 5																																																			
QY	2200	TTAGGAAGGACCCCTCAGGACACCCCTCTCTGCACCCCTGFGTCCCTCTAGAGTAGCTAGCTC	2259									Db	246	TTAGGAAGGACCCCTCAGGACACCCCTCTCTGCACCCCTGFGTCCCTCTAGAGTAGCTAGCTC	187									QY	2260	TGAGGCCCAAGGTGGGTACAAAGCAGGTATGGCCCTCAGAGATGCAGCCCTGCTGCTGGC	2319									Db	186	TGAGGCCCAAGGTGGGTACAAAGCAGGTATGGCCCTCAGAGATGCAGCCCTGCTGCTGGC	127									QY	2320	TTTTTCAGTCAGAGGTTGGGGGCTGGCCAGCCAGCTGCCTTTGCCCTGGCCGCTCTTACT	2379									Db	126	TTTTTCAGTCAGAGGTTGGGGGCTGGCCAGCCAGCTGCCTTTGCCCTGGCCGCTCTTACT	67									QY	2380	CCCTCCCTCTGCTCTCTCAGCTTCAGGTCCTATTTTTCACCTTTTCTTAAATAAAGAAATC	2439									Db	66	CCCTCCCTCTGCTCTCTCAGCTTCAGGTCCTATTTTTCACCTTTTCTTAAATAAAGAAATC	7									QY	2440	AG 2441										Db	6	AG 5																																																									
						Db	246	TTAGGAAGGACCCCTCAGGACACCCCTCTCTGCACCCCTGFGTCCCTCTAGAGTAGCTAGCTC	187									QY	2260	TGAGGCCCAAGGTGGGTACAAAGCAGGTATGGCCCTCAGAGATGCAGCCCTGCTGCTGGC	2319									Db	186	TGAGGCCCAAGGTGGGTACAAAGCAGGTATGGCCCTCAGAGATGCAGCCCTGCTGCTGGC	127									QY	2320	TTTTTCAGTCAGAGGTTGGGGGCTGGCCAGCCAGCTGCCTTTGCCCTGGCCGCTCTTACT	2379									Db	126	TTTTTCAGTCAGAGGTTGGGGGCTGGCCAGCCAGCTGCCTTTGCCCTGGCCGCTCTTACT	67									QY	2380	CCCTCCCTCTGCTCTCTCAGCTTCAGGTCCTATTTTTCACCTTTTCTTAAATAAAGAAATC	2439									Db	66	CCCTCCCTCTGCTCTCTCAGCTTCAGGTCCTATTTTTCACCTTTTCTTAAATAAAGAAATC	7									QY	2440	AG 2441										Db	6	AG 5																																																															
Db	246	TTAGGAAGGACCCCTCAGGACACCCCTCTCTGCACCCCTGFGTCCCTCTAGAGTAGCTAGCTC	187									QY	2260	TGAGGCCCAAGGTGGGTACAAAGCAGGTATGGCCCTCAGAGATGCAGCCCTGCTGCTGGC	2319									Db	186	TGAGGCCCAAGGTGGGTACAAAGCAGGTATGGCCCTCAGAGATGCAGCCCTGCTGCTGGC	127									QY	2320	TTTTTCAGTCAGAGGTTGGGGGCTGGCCAGCCAGCTGCCTTTGCCCTGGCCGCTCTTACT	2379									Db	126	TTTTTCAGTCAGAGGTTGGGGGCTGGCCAGCCAGCTGCCTTTGCCCTGGCCGCTCTTACT	67									QY	2380	CCCTCCCTCTGCTCTCTCAGCTTCAGGTCCTATTTTTCACCTTTTCTTAAATAAAGAAATC	2439									Db	66	CCCTCCCTCTGCTCTCTCAGCTTCAGGTCCTATTTTTCACCTTTTCTTAAATAAAGAAATC	7									QY	2440	AG 2441										Db	6	AG 5																																																																					
						QY	2260	TGAGGCCCAAGGTGGGTACAAAGCAGGTATGGCCCTCAGAGATGCAGCCCTGCTGCTGGC	2319									Db	186	TGAGGCCCAAGGTGGGTACAAAGCAGGTATGGCCCTCAGAGATGCAGCCCTGCTGCTGGC	127									QY	2320	TTTTTCAGTCAGAGGTTGGGGGCTGGCCAGCCAGCTGCCTTTGCCCTGGCCGCTCTTACT	2379									Db	126	TTTTTCAGTCAGAGGTTGGGGGCTGGCCAGCCAGCTGCCTTTGCCCTGGCCGCTCTTACT	67									QY	2380	CCCTCCCTCTGCTCTCTCAGCTTCAGGTCCTATTTTTCACCTTTTCTTAAATAAAGAAATC	2439									Db	66	CCCTCCCTCTGCTCTCTCAGCTTCAGGTCCTATTTTTCACCTTTTCTTAAATAAAGAAATC	7									QY	2440	AG 2441										Db	6	AG 5																																																																											
QY	2260	TGAGGCCCAAGGTGGGTACAAAGCAGGTATGGCCCTCAGAGATGCAGCCCTGCTGCTGGC	2319									Db	186	TGAGGCCCAAGGTGGGTACAAAGCAGGTATGGCCCTCAGAGATGCAGCCCTGCTGCTGGC	127									QY	2320	TTTTTCAGTCAGAGGTTGGGGGCTGGCCAGCCAGCTGCCTTTGCCCTGGCCGCTCTTACT	2379									Db	126	TTTTTCAGTCAGAGGTTGGGGGCTGGCCAGCCAGCTGCCTTTGCCCTGGCCGCTCTTACT	67									QY	2380	CCCTCCCTCTGCTCTCTCAGCTTCAGGTCCTATTTTTCACCTTTTCTTAAATAAAGAAATC	2439									Db	66	CCCTCCCTCTGCTCTCTCAGCTTCAGGTCCTATTTTTCACCTTTTCTTAAATAAAGAAATC	7									QY	2440	AG 2441										Db	6	AG 5																																																																																	
						Db	186	TGAGGCCCAAGGTGGGTACAAAGCAGGTATGGCCCTCAGAGATGCAGCCCTGCTGCTGGC	127									QY	2320	TTTTTCAGTCAGAGGTTGGGGGCTGGCCAGCCAGCTGCCTTTGCCCTGGCCGCTCTTACT	2379									Db	126	TTTTTCAGTCAGAGGTTGGGGGCTGGCCAGCCAGCTGCCTTTGCCCTGGCCGCTCTTACT	67									QY	2380	CCCTCCCTCTGCTCTCTCAGCTTCAGGTCCTATTTTTCACCTTTTCTTAAATAAAGAAATC	2439									Db	66	CCCTCCCTCTGCTCTCTCAGCTTCAGGTCCTATTTTTCACCTTTTCTTAAATAAAGAAATC	7									QY	2440	AG 2441										Db	6	AG 5																																																																																							
Db	186	TGAGGCCCAAGGTGGGTACAAAGCAGGTATGGCCCTCAGAGATGCAGCCCTGCTGCTGGC	127									QY	2320	TTTTTCAGTCAGAGGTTGGGGGCTGGCCAGCCAGCTGCCTTTGCCCTGGCCGCTCTTACT	2379									Db	126	TTTTTCAGTCAGAGGTTGGGGGCTGGCCAGCCAGCTGCCTTTGCCCTGGCCGCTCTTACT	67									QY	2380	CCCTCCCTCTGCTCTCTCAGCTTCAGGTCCTATTTTTCACCTTTTCTTAAATAAAGAAATC	2439									Db	66	CCCTCCCTCTGCTCTCTCAGCTTCAGGTCCTATTTTTCACCTTTTCTTAAATAAAGAAATC	7									QY	2440	AG 2441										Db	6	AG 5																																																																																													
						QY	2320	TTTTTCAGTCAGAGGTTGGGGGCTGGCCAGCCAGCTGCCTTTGCCCTGGCCGCTCTTACT	2379									Db	126	TTTTTCAGTCAGAGGTTGGGGGCTGGCCAGCCAGCTGCCTTTGCCCTGGCCGCTCTTACT	67									QY	2380	CCCTCCCTCTGCTCTCTCAGCTTCAGGTCCTATTTTTCACCTTTTCTTAAATAAAGAAATC	2439									Db	66	CCCTCCCTCTGCTCTCTCAGCTTCAGGTCCTATTTTTCACCTTTTCTTAAATAAAGAAATC	7									QY	2440	AG 2441										Db	6	AG 5																																																																																																			
QY	2320	TTTTTCAGTCAGAGGTTGGGGGCTGGCCAGCCAGCTGCCTTTGCCCTGGCCGCTCTTACT	2379									Db	126	TTTTTCAGTCAGAGGTTGGGGGCTGGCCAGCCAGCTGCCTTTGCCCTGGCCGCTCTTACT	67									QY	2380	CCCTCCCTCTGCTCTCTCAGCTTCAGGTCCTATTTTTCACCTTTTCTTAAATAAAGAAATC	2439									Db	66	CCCTCCCTCTGCTCTCTCAGCTTCAGGTCCTATTTTTCACCTTTTCTTAAATAAAGAAATC	7									QY	2440	AG 2441										Db	6	AG 5																																																																																																									
						Db	126	TTTTTCAGTCAGAGGTTGGGGGCTGGCCAGCCAGCTGCCTTTGCCCTGGCCGCTCTTACT	67									QY	2380	CCCTCCCTCTGCTCTCTCAGCTTCAGGTCCTATTTTTCACCTTTTCTTAAATAAAGAAATC	2439									Db	66	CCCTCCCTCTGCTCTCTCAGCTTCAGGTCCTATTTTTCACCTTTTCTTAAATAAAGAAATC	7									QY	2440	AG 2441										Db	6	AG 5																																																																																																															
Db	126	TTTTTCAGTCAGAGGTTGGGGGCTGGCCAGCCAGCTGCCTTTGCCCTGGCCGCTCTTACT	67									QY	2380	CCCTCCCTCTGCTCTCTCAGCTTCAGGTCCTATTTTTCACCTTTTCTTAAATAAAGAAATC	2439									Db	66	CCCTCCCTCTGCTCTCTCAGCTTCAGGTCCTATTTTTCACCTTTTCTTAAATAAAGAAATC	7									QY	2440	AG 2441										Db	6	AG 5																																																																																																																					
						QY	2380	CCCTCCCTCTGCTCTCTCAGCTTCAGGTCCTATTTTTCACCTTTTCTTAAATAAAGAAATC	2439									Db	66	CCCTCCCTCTGCTCTCTCAGCTTCAGGTCCTATTTTTCACCTTTTCTTAAATAAAGAAATC	7									QY	2440	AG 2441										Db	6	AG 5																																																																																																																											
QY	2380	CCCTCCCTCTGCTCTCTCAGCTTCAGGTCCTATTTTTCACCTTTTCTTAAATAAAGAAATC	2439									Db	66	CCCTCCCTCTGCTCTCTCAGCTTCAGGTCCTATTTTTCACCTTTTCTTAAATAAAGAAATC	7									QY	2440	AG 2441										Db	6	AG 5																																																																																																																																	
						Db	66	CCCTCCCTCTGCTCTCTCAGCTTCAGGTCCTATTTTTCACCTTTTCTTAAATAAAGAAATC	7									QY	2440	AG 2441										Db	6	AG 5																																																																																																																																							
Db	66	CCCTCCCTCTGCTCTCTCAGCTTCAGGTCCTATTTTTCACCTTTTCTTAAATAAAGAAATC	7									QY	2440	AG 2441										Db	6	AG 5																																																																																																																																													
						QY	2440	AG 2441										Db	6	AG 5																																																																																																																																																			
QY	2440	AG 2441										Db	6	AG 5																																																																																																																																																									
						Db	6	AG 5																																																																																																																																																															
Db	6	AG 5																																																																																																																																																																					

RESULT	9.			
LOCUS	AA639663/c	428 bp	mRNA	23-OCT-1997
DEFINITION	n83q12.s1 NCI_CGAP_Co9 Homo sapiens cDNA clone IMAGE:115882 3',		EST	
ACCESION	AA639663		mRNA sequence.	
VERSION	AA639663			
KEYWORDS	AA639663.1	GI:2563442		
SOURCE	EST.			
	human.			

ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 428)  
 AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 JOURNAL Unpublished (1997)  
 COMMENT On Jan 19, 1998 this sequence version replaced gi:2285718.  
 Contact: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550  
 Email: Robert.Strausberg@nih.gov  
 Tissue Procurement: Ilan Kirsch, M.D., Michael R. Emmert-Buck,  
 M.D., Ph.D.  
 CDNA Library Preparation: M. Bento Soares, Ph.D.  
 CDNA Library Arrayed by: Greg Lennon, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
[www-bio.llnl.gov/bbrp/image/image.html](http://www-bio.llnl.gov/bbrp/image/image.html)  
 Seq primer: -40ml3 fwd, ET from Amersham.  
 FEATURES  
 source Location/Qualifiers  
 1. .428  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:1158982"  
 /clone\_lib="NCI-CGAP\_Co9"  
 /tissue\_type="colon tumor RER+"  
 /lab\_host="DH10B"  
 /notes="Organ: colon; Vector: pT7T3D-Pac (Pharmacia) with a  
 modified polylinker; 1st strand cDNA was prepared from  
 RER+ colon tumor, and was then primed with a Not I -  
 oligo(dT) primer. Double-stranded cDNA was ligated to Eco  
 RI adaptors (Pharmacia), digested with Not I and cloned  
 into the Not I and Eco RI sites of the modified pT7T3  
 vector. Library is not normalized. Library was  
 constructed by Bento Soares and M. Fatima Ronaldo  
 (Soares4)."  
 101 a 101 c 145 g 81 t  
 BASE COUNT  
 ORIGIN

BASE COUNT	101 a	101 c	145 g	81 t	
ORIGIN	(bases=47)				
Query Match	16.7%; Score 409; DB 29; Length 428;				
Best Local Similarity	99.8%; Pred. No. 2e-90;				
Matches 420; Conservative	0; Mismatches 0; Indels 1; Gaps 1;				
QY	2021	GCAGCTGACATTTCAACCGCGCCCGCCAGGCTGTGCTGGGTCCCCCAGCCGCCACAGCCCTCA	2080		
Db	428	GCAGCTGACATTTCAACCGCGCCCGCCAGGCTGTGCTGGGTCCCCCAGCCGCCACAGCCCTCA	370		
QY	2081	CAGATPCCCATTTGCTCTCTGGCTTCTCCCAACCCCTAGGTGGGACAGTGAAGGGAGCA	2140		
Db	369	CAGATPCCCATTTGCTCTCTGGCTTCTCCCAACCCCTAGGTGGGACAGTGAAGGGAGCA	310		
QY	2141	GTTTAACCAAGAATTGCTGCTGCCCTTAGGGTCTCAGTCCCTCTCAGGAATCCCTCT	2200		
Db	309	GTTTAACCAAGAATTGCTGCTGCCCTTAGGGTCTCAGTCCCTCTCAGGAATCCCTCT	250		
QY	2201	TAGGAAGGACCCCTCAGGACACCCTCTCTGCACCCCTGTGGTCCCTCAGATAGCTAGCTCT	2260		
Db	249	TAGGAAGGACCCCTCAGGACACCCTCTCTGCACCCCTGTGGTCCCTCAGATAGCTAGCTCT	190		
QY	2261	GAGGCCCAAGTGGGTACAAAGCAGGTATGGCCCTCAGAGATGCAGCCTGCTGCTGGCT	2320		
Db	189	GAGGCCCAAGTGGGTACAAAGCAGGTATGGCCCTCAGAGATGCAGCCTGCTGCTGGCT	130		
QY	2321	TTTCAGTCAGAGGTTGGGGGCTGGCCAGCCAAAGCTGCCTTGCCCTGGCCGCTCTTTACTC	2380		
Db	129	TTTCAGTCAGAGGTTGGGGGCTGGCCAGCCAAAGCTGCCTTGCCCTGGCCGCTCTTTACTC	70		
QY	2381	CCTCCTCTGCTCTCACATTCAGGTCCATGTATTTACATTTTCTTAAATAAAGAATCA	2440		
Db	69	CCTCCTCTGCTCTCACATTCAGGTCCATGTATTTACATTTTCTTAAATAAAGAATCA	10		

```

QY 2441 G 2441
Db 9 G 9

RESULT 10
AL306592/c
DEFINITION mRNA sequence.
ACCESSION AL306592
VERSION AL306592.1 GI:3989663
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 410)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Apr 7, 1998 this sequence version replaced gi:3036105.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1350
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone Distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 693 Std Error: 0.00
Seq Primer: --40UP from Gibco
High quality sequence scop: 408.
Location/Qualifiers
1. .410
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1899142"
/clone_lib="NCI-CGAP_Lu5"
/tissue_type="carcinoid"
/lab_host="DH10B"
/note="Organ: lung; Vector: pTV3D-pac (Pharmacia) with a
modified polylinker; 1st strand cDNA was prepared from
neuroendocrine lung carcinoid, and was then primed with a
Not I - oligo(dT) primer. Double-stranded cDNA was ligated
to Eco RI adaptors (Pharmacia), digested with Not I and
cloned into the Not I and Eco RI sites of the modified
pTV3 vector. Library is normalized. Library was
constructed by Bento Soares and M. Fatima Bonaldo."
BASE COUNT 96 a 99 c 137 g 77 t 1 others
ORIGIN
Query Match 16.5%; Score 405.4; DB 37; Length 410;
Best Local Similarity 99.5%; Pred. No. 1.5e-89;
Matches 406; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2034 AACCCGGCCCCAGGTCCTGGTCCCGCCACCCACAGCCCTCACAGCATTCGCCAT 2093
Db 410 AACCCGGCCCCAGGTCCTGGTCCCGCCACCCACAGCCCTCACAGCATTCGCCAT 351

QY 2094 TGCTCTGCTCTTCCCGACCCCTAGTGGGACAGTGAAGGGAGGAGCTTTACAGAG 2153
Db 350 TGCTCTGCTCTTCCCGACCCCTAGTGGGACAGTGAAGGGAGGAGCTTTACAGAG 291

QY 2154 ATTGCTGCTGCCCTTAGGGTCTCAGCTCCCTCCTCAGGAATCCCTTTAGGAAGGACCT 2213
Db 290 ATTGCTGCTGCCCTTAGGGTCTCAGCTCCCTCCTCAGGAATCCCTTTAGGAAGGACCT 231

QY 2214 CAGGACACCCCTCTCTGCACCCTGTGGTCTCTAGAGTAGCTAGTCTAGGCCCCCAAGGT 2273
Db 230 CAGGACACCCCTCTCTGCACCCTGTGGTCTCTAGAGTAGCTAGTCTAGGCCCCCAAGGT 171

QY 2274 GGGTACAAGCAGGTATGGCCCTCAGAGATGAGAGCTGCTGCTGCTTTTTCAGTCAGAGG 2333
Db 170 GGGTACAAGCAGGTATGGCCCTCAGAGATGAGAGCTGCTGCTGCTTTTTCAGTCAGAGG 111

QY 2334 GTTGGGGCTGCCAGCCAGCTGCTTCCCTGGCGCTCTTACTCCCTCCCTCTGCTG 2393
Db 110 GTTGGGGCTGCCAGCCAGCTGCTTCCCTGGCGCTCTTACTCCCTCCCTCTGCTG 51

QY 2394 TCTCACTTCAGGTCCATGATTTTCACTTTTCTTAATAAAGAATCAG 2441
Db 50 TCTCACTTCAGGTCCATGATTTTCACTTTTCTTAATAAAGAATCAG 3

RESULT 11
AF046331 696 bp DNA GSS 21-APR-1998
LOCUS Mus musculus clone OST8023, genomic survey sequence.
DEFINITION Mus musculus clone OST8023, genomic survey sequence.
ACCESSION AF046331
VERSION AF046331.1 GI:3005202
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 696)
AUTHORS Zambrowicz,B.P., Friedrich,G.A., Buxton,E.C., Lilleberg,S.L.,
Person,C. and Sands,A.T.
TITLE Disruption and sequence identification of 2,000 genes in mouse
embryonic stem cells
JOURNAL Nature 392 (6676), 508-511 (1998)
MEDLINE 98219085
REFERENCE 2 (bases 1 to 696)
AUTHORS Zambrowicz,B.P., Friedrich,G.A., Buxton,E.C., Lilleberg,S.L.,
Person,C. and Sands,A.T.
TITLE Direct Submission
JOURNAL Submitted (06-FEB-1998) OmniBank, Lexicon Genetics Incorporated,
4000 Research Forest Drive, The Woodlands, TX 77381, USA
FEATURES
source
1. .696
/organism="Mus musculus"
/strain="129Sv/Ev"
/db_xref="taxon:10090"
/clone="OST8023"
/cell_type="embryonic stem cell"
/note="OmniBank Sequence Tag; exon trapped sequence"
BASE COUNT 166 a 193 c 185 g 131 t 21 others
ORIGIN
Query Match 16.5%; Score 405; DB 123; Length 696;
Best Local Similarity 82.5%; Pred. No. 2.1e-89;
Matches 490; Conservative 0; Mismatches 100; Indels 4; Gaps 3;

QY 704 TTCCCTAATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 763
Db 29 TTCCCTAATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 88

QY 764 CAGCTCAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 823
Db 89 CGGCCCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 148

QY 824 TCCAGTACGACGCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 883
Db 149 TCCAGTATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 208

QY 884 AGGAGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 943
Db 209 AGGAGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 268

```



and -minmatch 12 options.

PCR Primers  
FORWARD: AGGAACAGCTATGACCAT  
BACKWARD: GTTTCCTCCAGCTACGACG  
Plate: 104 row: I column: 9  
Seq primer: ATTTAGGTGACACTATAG.

# FEATURES

source  
1..539  
/organism="Bos taurus"  
/db\_xref="taxon:9913"  
/clone\_lib="MARC 1B0V"  
/tissue\_type="pooled"  
/lab\_host="DH10B"  
/note="vector: pcwv sport6; Site\_1: XbaI; Site\_2: XhoI;  
Library made from pooled tissue from lymph node, ovary,  
fat, hypothalamus, and pituitary."  
119 a 151 c 197 g 72 t

BASE COUNT  
ORIGIN

Query Match 16.4%; Score 402; DB 79; Length 539;  
Best Local Similarity 88.0%; Pred. No. 1.le-88;  
Matches 438; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

QY 121 AGGGGCTCTGCTGTACACAGAGCGGGCAAGGTCCCCAGGCCAGGATGTCCAGGCCTG 180

Db 42 AGCGGACAGCAGACAGCTGGCGCCAGGCAAGGTCCCGGCGCAGGATGTGGGCGTG 101

QY 181 GTCTTGGGCGAGCGGATGAGCTCGAGGCCACCGCTCAGCCAGAGGAGATCCTGGGG 240

Db 102 GTGTGGGACAGGGGATGAGCTCGAGGCCACCGCTCAGCCAGGAGGATCCTGGGG 161

QY 241 AGCACAGCGTGTGTAGCCAGAGGCTAGAGGCCCTTACGCTGACAGGAGGAGGAGGCTG 300

Db 162 AGCACCGTGTGTAGCCAGAGGCTAGAGGCCCTTACGCTGACAGGAGGAGGAGGCTG 221

QY 301 CAAGCTGTCCAGACCATTCAGTGTCTGCAGCAGGAGGCCATGAGAGGCGCTGGTG 360

Db 222 CAAGCTGTCCAGACCATTCAGTGTCTGCAGCAGGAGGCCATGAGAGGCGCTGGTA 281

QY 361 CATGAGAAGCGCGGCGAGCTCGCGGTTCTATGAAACATTTAGCTCGGCTGAGTGAG 420

Db 282 CATGAGAAGCGCGGCGAGCTCGCGGTTCTATGAAACATTTAGCTCGGCTGAGTGAG 341

QY 421 GCCAGGTGATGCTGCTCTACAGCAGCCAGCTGAGCAGTGGAGTGGAGAAACAGAG 480

Db 342 GCCAGGTGATGCTGCTTGGCCCAACCCAGCTGAGCAGTGGAGTGGAGAAACAGAG 401

QY 481 CTGGGCTCAGTGGCGGCGCTATGCCAGAGAACCACTGCGGATGAGCTGGCT 540

Db 402 CTGGGCTCAGTGGCGGCGCTATGCCAGAGAACCACTGCGGATGAGCTGGCT 461

QY 541 GGCACCCAGCGGCTACAGCGAGTGAACAGCTGTGGCTCAGCTGAGGAGGAGAAAG 600

Db 462 GGCACCCAGCGGCTACAGCGAGTGAACAGCTGTGGCTCAGCTGAGGAGGAGAAAG 521

QY 601 AAGCAGCTGGAGTTCCTG 618

Db 522 AAGCAGCTGGAGTTCCTG 539

RESULT 14  
AW146303/c 767 bp mrna EST 03-APR-2000  
LOCUS um54d12.x1 Sugano mouse embryo mewa Mus musculus cDNA clone  
DEFINITION IMAGE:2259095.3' similar to FR:088448 088448 KINESIN LIGHT CHAIN 2.  
// mRNA sequence.

ACCESSION AW146303  
VERSION AW146303.1 GI:6168039  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

# REFERENCE

AUTHORS  
JOURNAL  
COMMENT

TITLE  
JOURNAL  
COMMENT

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 767)  
Marta, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T.,  
Underwood, K., Theising, B., Allen, M., Bowers, Y.,  
Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R.,  
Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R.,  
Waterson, R. and Wilson, R.  
The WashU-NCI Mouse EST Project 1999

Unpublished (1999)  
On Apr 7, 1998 this sequence version replaced gi:3035979.

Contact: Maria M/WashU-NCI Mouse EST Project 1999  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810

Email: mouseest@wustl.wustl.edu  
This clone is available royalty-free through LLNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Seq primer: custom primer used  
High quality sequence stop: 511.

# FEATURES

source  
1..767  
/organism="Mus musculus"  
/strain="C57BL"  
/db\_xref="taxon:10090"  
/clone="IMAGE:2259095"  
/clone\_lib="Sugano mouse embryo mewa"  
/dev\_stage="embryo, 14 dpc"  
/lab\_host="DH10B"  
/note="vector: PME18S-FL3; Site\_1: DraIII (CACTGTGTG);  
Site\_2: DraIII (CACCATGTG); 1st strand cDNA was primed  
with an oligo(dT) primer [ATGGCCCTTTTTTTTTTTT];  
double-stranded cDNA was ligated to a DraIII adaptor  
[TGTGGCTACTGG], digested and cloned into distinct DraIII  
sites of the PME18S-FL3 vector (5' site CACTGTGTG, 3' site  
CACCATGTG). XhoI should be used to isolate the cDNA  
insert. Size selection was performed to exclude fragments  
<1.5kb. Library constructed by Dr. Sumio Sugano  
(University of Tokyo Institute of Medical Science).  
Custom primers for sequencing: 5' end primer  
CTTCTGCTTAAAGCTGGC and 3' end primer  
CGACTGCGCTCGAGCACA."

BASE COUNT 170 a 226 c 213 g 153 t 5 others  
ORIGIN

Query Match 16.3%; Score 401; DB 64; Length 767;  
Best Local Similarity 77.0%; Pred. No. 2.le-88;  
Matches 608; Conservative 0; Mismatches 155; Indels 27; Gaps 9;

QY 1654 CAGGGCACTGACCCCTATCAGCCAGACGAGAGTGGCAGAGCTGCTTGGGGAGATGATGTT 1713

Db 767 CAGGGTACTGCCCATCAAGTCAGACACAGGCTGGCAGAGTGCCTCGGGGAAGGAGATGTT 708

QY 1714 AGAAGGACCTCCAGAGGCGCCCTGGAGACAGTGTGAAATTCAGAGGCTGCTGAAGATGCT 1773

Db 707 AGAAGGCCATCCAGAGGCTCTCGAGACAGCTGAAGTTCAGAGNAGCTGAACGCA 648

QY 1774 TCTGTGGCTGTGGATGTGTCGGGGATGGCAGTGGGACCCCTGCAGAGAGTGTCTCTT 1833

Db 647 TCAGTGGCTGTGGATGTGTCAGNNGATGGCAGTGGCACCCTCCAGAGGANNCGCTCTCTG 588

QY 1834 GGCAAGATCCGGATGTGCTCCGCAAGACAGTGAACCTCTTGTGTGAGAAAGCTCCAGGGG 1893

Db 587 GGCAAGATCCGGATGTGCTTCGTAGGACAGTGAAGTGTGAGAAAGCTCCAGGGG 528

QY 1894 ACTGAGGCTCGGCCCTCCAGACAGCAACATGAAGCAGCAGCCTCTTGAACATATCTGAAC 1953

Db 527 ACGGAGCCGCGACCTTCCAGCAGCAGCATGAAGCGGGCGGCTCGCTGAACACTACTTAAAC 468

QY 1954 CAACCTAGTGCAGCAGCCCTCCAGGCTCTCCCGGGGCTCAGTGCAGCAGCATGAGGACCTC 2013

Db 467 CAACCCATGCAGCCCTCCAGGCTCTCCCGGGGCTCAGCAGCAGCATGAGGACCTC 408



---